DOCUMENT RESUME

ED 181 443 CS 005 228

TITLE Evaluation of Outcomes, 1977-78: An Evaluation Sytem Report on Reading Programs and Reading Achievement: Part II-A Technical Report.

INSTITUTION District of Columbia Public Schools, Washington, D.C. Dept. of Research and Evaluation.

PUB DATE Nov 79

NOTE 224p.; For related documents see ED 169 491, ED 169 492, and ED 178 883. Tables A, B, and C may be marginally legible.

EDRS PRICE MF01/PC09 Plus Postage.

DESCRIPTORS Elementary Education: *Program Evaluation: *Reading Achievement: *Reading Interests: *Reading Programs: *Reading Research: *Teacher Behavior

IDENTIFIERS *District of Columbia

ABSTRACT

This report presents results of analyses of factors affecting reading achievement of 14,318 pupils in 34 elementary schools in Washington, D.C., as measured by the percentage of objectives mastered on the Prescriptive Reading Test and students' interest in reading for the school year 1977-78. It also presents descriptive information about students, classes, and reading programs for that year. The data in this report were collected using three basic instruments: the Classroom/Student Background Form, the Prescriptive Reading Test, and the Reading Program Form. Findings of the report are given in three parts: reading achievement and testing, interest in reading, and evaluations of selected types of reading programs. (TJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181443

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

District of Columbia Public Schools

EVALUATION OF OUTCOMES, 1977-78: AN EVALUATION SYSTEM REPORT ON READING PROGRAMS AND READING ACHIEVEMENT

PART II - A TECHNICAL REPORT

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
District of Columbia Public Schools

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Vincent E. Reed
Superintendent of Schools
November 1979

by
Division of Research and Evaluation
November 1979

CS005227

District of Columbia Public Schools

OFFICE OF SUPERINTENDENT

Superintendent of Schools	Vincent E. Reed
Vice Superintendent of Schools	Elizabeth C. Yancey

OFFICE FOR EDUCATIONAL ACCOUNTABILITY

Associate Superintendent	Wilbur A. Millard

DIVISION OF RESEARCH AND EVALUATION

Assistant Superintendent	Mildred P. Cooper
Director, Evaluation System	Herman Cobb, Jr.
Assistant Director	Beulah Glenn

Data Collectors (1977-78)

Dorothy Anderson	Iris Proctor
Madeline Baker	Carlos Reeder
Linda Clifford	William Seldon
Constance Dickerson	Olivia Simons
Warren Kullberg	Debra White

ACKNOWLEDGMENTS

The Division of Research and Evaluation extends appreciation to the many individuals who have contributed both in general and in terms of this report to the Evaluation System of the Public Schools of the District of Columbia. Deep gratitude is expressed to the teachers, reading specialists, students, administrators, and others in the Evaluation System schools for their continuing cooperation and interest. We trust that they will find in the two parts of the report (the summary and the technical report) an indication of the kinds of information and evaluations that their efforts have made and will continue to make possible. We also thank the Division of Automated Information Systems (DAIS) which has handled the preparation and merging of computer tapes and those consultants--Dr. H. Russell Cort, Jr., Dr. Barry Smith, Dr. Walton Smith, and Bruce Jonas who provided major assistance in the preparation of this report.

FEB 2 6 1980

PREFACE

This report is the third in a series of three Evaluation System reports to be issued by the Division of Research and Evaluation (DRE) prior to the 1979-80 school year.

This report provides an analysis of data from school year 1977-78.

The first two reports presented analyses of data from school years 1975-76 and 1976-77.

Each report is in two parts: "Part I - A Summary" and "Part II - A Technical Report."

Additional reports will be developed and issued upon completion of further analysis of the data.

TABLE OF CONTENTS

	Page
LIST OF TABLES	VIII
LIST OF FIGURES	IX
CHAPTER	
I. INTRODUCTION	1
II. PURPOSE AND SCOPE	3
Objectives	3
The Data Collection and the Data	3
Limitations	6
Definitions	8
III. METHODOLOGY	9
Data Collection	9
The Data Bank	9
Future Developments in the Evaluation System	11
IV. CHARACTERISTICS OF THE 1977-78 CLASSES	14
Purpose of This Chapter	14
The Variables Included	14
Class-Level Descriptive Statistics	16
Characteristics of Reading Programs	33
V. RESULTS BASED ON REGRESSION ANALYSES	44
The Regression Analyses	44
Regression on Posttest Performance Scores	51
Regression Analysis of Reading Interest	66
VI. EVALUATIONS OF READING PROGRAMS	85
Methods	85
Comparison of the First Grade Reading Programs	88
Comparisons of Reading Programs at the Second Grade Level	97
Comparisons of Reading Programs at the Third Grade Level	111
Comparisons of Reading Programs at the Fourth Grade Level	120
Comparisons of Reading Programs at the Fifth Grade Level	128
Comparisons of Reading Programs at the Sixth Grade Level	134
General	140
Summary	144

TABLE OF CONTENTS (continued)

VII. CONCLUSIONS..... 155

Reading Achievement and Testing..... 155
Interest in Reading..... 156
Evaluations of Selected Types of Reading Programs..... 157

APPENDICES

A. REPLICA OF THE CLASSROOM/STUDENT BACKGROUND FORM.. 161

B. REPLICA OF THE READING PROGRAM FORM..... 164

C. EVALUATION SYSTEM SCHOOLS..... 169

D. CORRELATION MATRIX ACROSS ALL GRADES REGRESSION ON POSTTEST SCORES: TABLE A..... 171

E. CORRELATION MATRIX FOR PRIMARY GRADES. TABLE B... 173

F. CORRELATION MATRIX FOR INTERMEDIATE GRADES: TABLE C..... 175

G. NUMBER OF STUDENTS AND CLASSES IN 34 EVALUATION SYSTEM SCHOOLS BY YEAR AND GRADE: TABLE D..... 177

LIST OF TABLES

Table		Page
I.	Characteristics of the 1977-78 Evaluation System Classes	17
II.	Benchmark Characteristics of Classes by Grade Level	30
III.	Percentages of Students in Different Reading Programs, by Grade and Year	35
IV.	Posttest Regression Summary: All Grades	53
V.	Posttest Score Regression - All Grades: Contribution of Sets	54
VI.	Posttest Regression Summary: Primary Grades	59
VII.	Posttest Score Regression - Primary Grades: Contribution of Sets	60
VIII.	Posttest Regression Summary: Intermediate Grades	64
IX.	Posttest Score Regression - Intermediate Grades: Contribution of Sets	65
X.	Reading Interest Regression Summary: All Grades	68
XI.	Reading Interest Score Regression - All Grades: Contribution of Sets	69
XII.	Reading Interest Regression Summary: Primary Grades	76
XIII.	Reading Interest Score Regression - Primary Grades: Contribution of Sets	77
XIV.	Reading Interest Regression Summary: Intermediate Grades	80
XV.	Reading Interest Score Regression - Intermediate Grades: Contribution of Sets	81
XVI.	Statistical Data for Each Reading Program Compared, Grade 1	89
XVII.	Comparison of Reading Programs by Receiving Extra Help, Grade 1	90
XVIII.	Composition of Samples for Analyses of Reading Programs, Grade 1	92

LIST OF TABLES (continued)

		Page
XIX.	Statistical Data for Each Reading Program Compared, Grade 2	98
XX.	Comparison of Reading Programs by Receiving Extra Help, Grade 2	101
XXI.	Composition of Samples for Analyses of Reading Programs, Grade 2	103
XXII.	Statistical Data for Each Reading Program Compared, Grade 3	112
XXIII.	Comparison of Reading Programs by Receiving Extra Help, Grade 3	114
XXIV.	Compositions of Samples for Analyses of Reading Program, Grade 3	115
XXV.	Statistical Data for Each Reading Program Compared, Grade 4	122
XXVI.	Comparison of Reading Programs by Receiving Extra Help, Grade 4	123
XXVII.	Composition of Samples for Analyses of Reading Programs, Grade 4	124
XXVIII.	Statistical Data for Each Reading Program Compared, Grade 5	129
XXIX.	Comparison of Reading Programs by Receiving Extra Help, Grade 5	131
XXX.	Composition of Samples for Analyses of Reading Programs, Grade 5	132
XXXI.	Statistical Data for Each Reading Program Compared, Grade 6	135
XXXII.	Comparison of Reading Programs by Receiving Extra Help, Grade 6	136
XXXIII.	Composition of Samples for Analyses of Reading Programs, Grade 6	138
XXXIV.	Types of Reading Programs Evaluated, 1975-76, 1976-77, and 1977-78, By Grade	145

LIST OF FIGURES

Figure		Page
1.	Regions of Significant Difference Between Pairs of Reading Programs, Grade 2	102
2.	Regions of Significant Difference Between Pairs of Reading Programs, Grade 3	116
3.	Regions of Significant Difference Between Pairs of Reading Programs, Grade 4	125

I. INTRODUCTION

The Evaluation System of the Public Schools of the District of Columbia has several purposes. At the most general level it is designed to enable the school system to evaluate the effectiveness of programs and to identify characteristics of programs that seem to increase effectiveness. Also, it is designed to enable the D.C. Public Schools to follow particular students over a period of years, to examine long-term effects of different programs, and to measure changes in program characteristics from year to year. The design of the Evaluation System is such that analyses or evaluations can be done at a number of levels: the individual student, the classroom, the school building, the grade level, the program, the region, and the overall school system.

An essential feature of the Evaluation System is its ability to correlate data. Thus, the System is designed to combine test data with program data, and to combine both of these with a variety of other data or measures such as: student absence, sex, age, size of class, student turnover rate for classes, type of class, involvement of aides, and many other such measures. These data or measures come from various sources. Combining them has been a major undertaking. This report presents various student, class, teacher, and program data combined with performance measures; in this case, test data.

The Evaluation System is designed to support the investigation of an almost limitless number of questions. It is thus the intention of the Division of Research and Evaluation, which developed and implements the

Evaluation System, that users of its reports (administrators, teachers, and others) become familiar with the System and its potential. The objective findings of the Evaluation System will provide decision-makers at all levels with sound information upon which to make judgments for future directions.

II. PURPOSE AND SCOPE

A. Objectives

This report presents results of analyses of factors affecting reading achievement as measured by the percentage of objectives mastered on the Prescriptive Reading Test (PRT), and students' interest in reading for the year 1977-78. It also presents descriptive information about students, classes, and reading programs for that year. Data for grades 1-6 are included in the report.

Although this is the third report to exhibit Evaluation System data for which such analyses have been made, the results reported here should be viewed as indicative. In some respects the data of this report are baseline data against which data for subsequent years can be compared. Not all the data collected or measures computed have been used in the analyses of this report, nor was it the intention to do so. Much of the data have already been issued to the Evaluation System schools in earlier reports.

B. The Data Collection and the Data

1. Collection Procedures

The data in this report was collected using three basic instruments:

a. The Classroom/Student Background Form (SBF) is a two-part machine-readable form designed especially for the Evaluation System by the Division of Research and Evaluation (DRE). One part of the form is for the collection of background data on the classroom and the other part is for the collection of background data on the students

in the class. This form is completed by staff from DRE in the fall and spring of the school year. See Appendix A.

b. The Prescriptive Reading Test (PRT), the school system's own criterion-referenced reading test, is administered in the fall and spring of each year. The test is constructed by levels ranging from A through E and G which cover the reading grade levels objectives for grades 1-6. Students are administered test levels according to their instructional level. Fall and spring test results were matched for each student by DRE.

c. The Reading Program Form (RP) is a two-part machine-readable form designed by DRE and completed by teachers during the month of May of each year. One part of the form is for the collection of classroom reading data and the other part is for the collection of individual student reading program data. See Appendix B.

Data was collected from the 34 elementary schools participating in the Evaluation System. The official October, 1977 total membership in these schools for grades 1-6 was 14,318 students. See Appendix C for the distribution of schools by region.

2. Test Data

The test data are percentages of reading objectives mastered in the fall of 1977 (pretest) and in the spring of 1978 (posttest) on the Prescriptive Reading Test (PRT), Levels A through E and G. Only the scores for students who had both pre and post scores for the same level of test were used.

Since many teachers administered different levels of the PRT to different students in their classes (as they should have), test raw scores were converted to percentages in order to arrive at a class average, or in order to make the range and units of test scores reasonably comparable for different students in the same reading program. The "raw score" for a test was the number of objectives mastered according to PRT criteria. The percentage computed was the number of objectives mastered divided by the total number of objectives in the test for the level administered. The number of objectives varied from 21 for Level A to 45 for Level B. Subsequent reports will examine subtest scores, computed in similar fashion.

3. Student Characteristics

Data, other than test scores, which pertain to individual students include sex, age, absences, total time enrolled in a class, student interest in reading, need for extra help in reading, how often student reads on own and student's class and ID number.

4. Teacher Characteristics

Characteristics of teachers include absences, reasons for using particular reading materials, freedom to choose own reading skills objectives to be taught, extent to which teachers coordinated instructional reading program with other school personnel and length of time, and period of time during the year the teacher was associated with a particular class.

5. Reading Programs

A description of each student's reading program in terms of materials, grade levels of materials, and relative emphasis given to each material was obtained from the reading form described in Item 1.c. While the reading program is a characteristic of each student, it is listed separately here as a reminder that a focus of the Evaluation System is programs. The programs, of course, may differ even for students within the same classroom.

6. Classroom Characteristics

Division of Research and Evaluation staff collected data pertaining to the classroom as a whole, not specifically to individual students (e.g., number of students eligible for free lunch program, number of students referred for various reasons, type of teaching situation, presence of an aide, availability of reading or mathematics specialists, and the like). In addition, a number of measures pertaining to the classroom as an aggregate were computed from data based on individuals (e.g., student absence rate, student turnover rate, class average percentage of objectives mastered, class average interest in reading, and class average need for extra help in reading).

C. Limitations

It is important that the reader understand the limitations on the generalizability of the results and conclusions reported here. First, there were large losses for some variables due to missing data. For

example, on test data given to the Division of Research and Evaluation the designation of test level was missing from the records of 7,072 out of 15,870 students. That was a loss of 45% of the cases for any analysis using test scores. Without knowledge of the level of test administered it was impossible to compute percentage of objectives mastered.[1/] Second, for purposes of this report, only data for students who were administered the same level of test, pre and post, were used. That automatically excluded all students who were administered a higher or lower level test in the fall (pre) than they were in the spring, (post). Third, there were losses of data for particular variables owing to the procedures used for computer tape files. Fourth, there were losses of data for particular variables owing to the failure of some forms to be filled out accurately or completely, but also owing to errors in optical scan processing of data from forms. The limitations described are mainly the consequences of limitations of resources, processing and the like, not of design limitations of the Evaluation System. Moreover, data loss has been considerably reduced from previous years.

In a data gathering program of this scope, it was unfeasible to evaluate the consistency of ratings of students made by teachers. That is, it was not feasible to do a form of test-retest. By the same token, it was not feasible to observe reading programs directly. All such reading data are based on the subjective reports of teachers.

In general, the data reported here are of inherent interest, but they must be viewed as preliminary. The analyses should be viewed

---

1/ A determination of test level for levels A and B can be made by counting the number of test items. Levels C, D, E, and G, however, all have 120 items.

as illustrative of some of the kinds of inquiries that can be made with the matched data of the Evaluation System.

The findings reported in the Summary (Report One, Part I) are stated in non-technical terms without detailed statistical back-up. Detailed statistical analyses, as well as technical discussions, are reported in the Technical Report (Report One, Part II) upon which the Summary Report is based.

D. Definitions

Classroom Characteristics - Include grade, room number, number of free lunch pupils, class size, and percent of male and female.

Learning Barriers - Those conditions which may confront a student and hinder the learning process, i.e., poor attendance, lack of proper nutrition, economic conditions, and medical/physical health.

Posttest Scores - The percentage of reading objectives mastered in the spring based on the total number of objectives possible at the test level administered.

PMT - Prescriptive Mathematics Test

Pretest Scores - The percentage of reading objectives mastered in the fall based on the total number of objectives possible at the test level administered.

PRT - Prescriptive Reading Test

Reading Behaviors - Those behaviors a student exhibits toward reading, i.e., amount of interest in reading and the degree to which student reads on his own.

Teacher Characteristics - Those descriptors of the teacher such as the day the teacher started with the class, the last day with the class, and teacher turnover rate.

III. METHODOLOGY

A. Data Collection

Since the present report is intended to parallel the earlier technical reports of the 1975-76 and 1976-1977 data, the basic approach to data collection, briefly described in Chapter II, was the same as for the earlier year.

Data were collected in the form of reading pretest and posttest scores, and test results for each student are available both in a composite score and objective by objective. In addition, the forms completed by teachers and Division of Research and Evaluation staff members provided a variety of important information concerning individual students and the classroom in which they were taught. The specific variables and sets of variables on which the data were collected are described in more detail in Chapter IV.

Once the forms had been completed, they were edited, scanned, and stored on magnetic tape. The test scores were also provided for both fall and spring and stored on tape.

B. The Data Bank

In an earlier technical report, we described the development of the Evaluation Data Acquisition, Control, and Analysis (EDACA) System as a component of the Evaluation System. As was noted earlier in the report, we are dealing, even in the pilot study, with data from some 14,318 students forming 562 classes in grades 1-6 in 34 schools. A computer-based data acquisition, storage, and processing system like EDACA is essential to the handling of such large quantities of data.

Using both the 1975-76 data base and the 1976-77 data base in part for the purpose of testing the EDACA System, it has been possible to make considerable progress on that system. The evolving concept is that of an efficient computer-based system that will be highly standardized, yet flexible enough to permit the incorporation and deletion of variables and the addition of larger numbers of schools, classes, and students, eventually to include the entire school system if desirable.

Computer Programs. The EDACA system involves a number of different computer programs, each fulfilling a specific need. Basically, however, there are three types of programs involved in the system: (a) manipulation programs permit the programmer to store and access the data files as well as to merge two or more files into a single file and to create new files containing partial or specialized information needed for particular analyses; (b) descriptive statistical programs provide information concerning frequencies, means (averages), and other basic statistics which help us to describe and thus better understand the nature of the classes and students appearing in our samples. These programs thus tell us, for example, how many classes from the total sample of classes appear in each of the six grades, how many classes have experienced a change of teacher during the school year, the average class size, the average number of class absences, and the average percentage of male students in the classes (out put from these descriptive programs is dealt with in some detail in Chapter IV); and (c) analytical programs permit various statistical analyses of the data to be undertaken. The two principal types of analysis applied to 1975-76, 1976-77, and 1977-78 data are regression analysis and analysis of covariance. These are described in two later chapters.

C. Future Developments in the Evaluation System

The EDACA System has been further developed beyond the point at which it stood when the second study was completed. It has now been partially standardized to provide the kinds of file manipulation, statistical description, and analysis included in both the present and earlier reports. The system is now being expanded to readily permit further types of analysis and thus permit us to learn even more from the data bases provided by the Evaluation System as it currently exists. Plans for the further development and modification of the EDACA System still involve the same major goals as stated in the earlier technical report:

1. Flexibility

The system is now flexible enough that it will permit reasonable additions and deletions of variables as new questions are raised and old ones are satisfactorily answered. Further work remains to fully document the flexibility of the system and to further develop that flexibility.

2. Expansion Capability

Although the Evaluation System does currently involve a large number of students, classes, and schools, it is recognized that this still represents a pilot stage in its development. Future plans call for the extension of the system to a larger number of elementary schools, to secondary schools, and eventually to the entire school system. This means that the capability for expansion in numbers of schools, classes, and students must be built into the developing EDACA system. It now seems clear that quantitative expansion of the computer analysis system will be straightforward. However, it is essential to recognize that some kinds of expansion may necessitate the continued use of a large scale computer system, such as that currently being employed.

3. Reduction of Data Loss

Data loss has been considerably reduced in the present study as compared with the 1975-76 and 1976-77 studies. Through further refinement of data processing procedures, it has been possible to identify 458 classes valid for analysis in the 1977-78 year. This represents an increase over the 396 valid classes in 1976-77 and the 229 in 1975-76. Further work is being done to produce an even greater reduction in data loss in future years.

4. Individualized Feedback

The present report focuses on data descriptions and analyses which cut across all available schools, classes, and students. This level of data processing is essential in providing important information concerning adjustments and modifications that can be made by the school system in its ongoing efforts to improve the reading and other skills of the individual child. It would also be desirable, however, to provide direct feedback to teachers, principals, and regional and central administrative personnel concerning their specific classes, schools, or regions. Thus, it would be helpful to the individual teacher to know in detail the characteristics of his or her classroom and the individual students in the classroom and to know how these characteristics compare with those generally characterizing similar classrooms in the school system. Such individualized profiles are not yet available, but it is hoped that the EDACA system can be expanded to include such a profiling procedure.

5. Rapid Turnaround

An important goal in the further development of the EDACA system is to provide rapid processing of data, such that feedback can be provided within a relatively short time following the collection of data. For

example, it would be desirable to provide feedback early in the fall semester concerning data collected the previous spring. This has become one of the major goals of the Evaluation System, and EDACA represents, in part, an effort to achieve this goal. Over the course of the three studies completed to date, turnaround time has been considerably reduced from nearly 2½ years for the 1975-76 data to almost 2 years for the 1976-77 data to just over 1 year for the 1977-78 data. It is expected that gradual further reduction in turnaround time is a reasonable goal for the future.

IV. CHARACTERISTICS OF THE 1977-78 CLASSES

A. Purpose of This Chapter

This chapter presents descriptive statistics pertaining to the 1977-78 Evaluation System schools. The data presented are for the samples used in major analyses (Chapters V & VI), with a few exceptions to be noted in the text. Descriptive statistics are presented for both classes and students as the units of analysis. Thus, the reader will find that there may be differences in the values of a particular variable for those two conditions. It is always necessary to keep in mind whether one is considering classes or students when considering the data.

B. The Variables Included

The descriptive statistics to be presented here are for variables that were used in analyses of relationships and differences, as noted above. For class-level fixed-order regression analyses of achievement, a total of 22 variables (21 predictors and 1 criterion variable) were used, including pretest and posttest measures. Regression analyses of interest in Reading used one additional predictor.[2/]

For student-level analyses of covariance, 6 covariates were selected. These were variables for which the measures pertained to the individual student.

To maintain sample sizes of classes in regression analyses, regressions were made for primary level samples of classes (grades 1-3) and for intermediate level samples of classes (grades 4-6). In effect, primary

---

2/ Given sets of analyses always used a list-wise deletion procedure to determine samples. Any case for which any variable was missing was thus excluded. Under that procedure losses of cases can rise rapidly as more variables are included if there are missing data. Thus, it is important to minimize the number of variables used.

grade classes were analyzed as a group, and intermediate grade classes were analyzed separately as another group. Thus, the descriptive statistics presented in this chapter pertain to these two main groups of classes as well as to the total group (grades 1-6). Multi-grade classes were included in these groups by taking the unweighted average grade of each multi-grade class. Thus, a 4-5 combination class would have a computed grade level of 4.5. A 4-5-6 multi-grade class would be assigned a grade level of 5. With the exception of certain analyses (which will be noted), class-level statistics are for sets as described, not for each grade level separately. Given this method of assigning grade levels to individual classes, the actual upper boundary of primary classes was 3.4. The lower boundary of intermediate classes was 3.5. The boundaries of all classes were, of course, 1-6.

Analyses of covariance of reading programs were done on a grade by grade basis. The analyses of covariance used the student as the unit of analysis. In one set of analyses, only students from single grade level classes were included. In a second set of analyses, students from multi-grade classes were included. The basis for assigning students from multi-grade classes to a particular grade level (1, 2, 3...6) for purposes of analysis was to define the lower and the upper boundary as $\pm$.5. Thus, a student from a multi-grade class with an assigned grade level of 3.4 would be included in an analysis of covariance of 3rd grade students. That would be the case even if the student literally was in the 4th grade part of the class. The reason for this is that grade level code pertains to the class in the Evaluation System, not to the student within the class.

C. Class-level Descriptive Statistics

The total analysis pool of classes for the 34 elementary schools in the Evaluation System in 1977-78 was 548.[3/] That is an average of 16 classes per school, or an average of 91 per grade level if only 6 grades are considered. The maximum useable samples were:

- For all grades combined: 458 classes
- For primary grades: 244
- For intermediate grades: 214

Thus, the useable samples of 458 classes were about 84% of the total pool.[4/]

Table I gives descriptive statistics for Primary grades (1-3), intermediate Grades (4-6), and the Total Useable Sample for all grades (1-6). The data for the three samples are shown in the same table to make it easy to compare the three groups on each variable. One should remember, however, that Primary classes actually range from grades 1-3.4, Intermediate ones from grades 3.5-6. Definitions of variables follow in the text below. The reader is reminded that the data in Table I refer to classes as units, not to individual students as units.

1. Definitions of Variables in Table I

Variables in Table I are defined below. The variables are lettered here as they are in Table I to aid the reader.

3/ 548 was not the total number of classes in the Evaluation System. It was the total number meeting certain ID matching criteria and an exclusion criterion. The latter was that any class for which the average membership of students in the class was less than 100 days was excluded. The total number of classes in the Evaluation System was 562 for 1977-78.

4/ The total analysis pool of 548 classes was reduced to 458 useable classes through a list-wise deletion process. That is, if any student in a class was missing data on any one of the selected variables the whole class was lost. Only students with pre-post matched data on the same level PRT were used.

TABLE I

CHARACTERISTICS OF THE 1977-78 EVALUATION SYSTEM CLASSES

Name of Variable	Level*	# of Classes	Mean	SD	Correlation With Posttest**	Correlation With Interest
a. Pretest % Objectives Mastered	Primary	244	50.32	11.14	.53	-.10
	Inter.	214	44.20	11.41	.51	-.16
	Total	458	47.33	11.67	.56	-.14
b. Posttest % Objectives Mastered	Primary	244	79.69	9.70	---	-.33
	Inter.	214	72.51	9.43	---	-.22
	Total	458	76.19	10.21	---	-.28
c. Grade (Ave. Grade Level)	Primary	244	1.92	.82	-.18	.00
	Inter.	214	4.89	.83	-.08	.08
	Total	458	3.31	1.70	-.39	.07
d. Test Level Count (Ave. # Different Test Levels Class)	Primary	244	1.76	.78	-.15	-.06
	Inter.	214	2.57	.99	.02	.09
	Total	458	2.14	.97	-.21	.04
e. Class Size	Primary	244	24.64	3.72	-.00	-.08
	Inter.	214	25.74	3.70	.01	-.06
	Total	458	25.16	3.74	-.05	.06
f. % Males (Class %)	Primary	244	51.46	9.01	.07	-.00
	Inter.	214	49.26	8.85	.02	.20
	Total	458	50.43	8.99	.09	-.09

*Note : Primary = Primary Grades 1 Through 3.4.
Inter. = Intermediate Grades 3.5 Through 6.
Total = Total Useable Sample

**Correlations of the variable designated with posttest are specifically with arcsine transformations of posttest class average percentages of objectives mastered.

TABLE I CONTINUED

Name of Variable	Level*	# of Classes	Mean	SD	Correlation With: Posttest**	Interest
g. Class Holding Rate (100-Turnover Rate)	Primary	224	88.72	7.98	.03	-.01
	Inter.	214	90.90	7.11	.08	-.08
	Total	458	89.74	7.65	-.01	-.03
h. Teacher Turnover Proportion	Primary	244	.02	.16	-.02	.18
	Inter.	214	.02	.14	-.04	.14
	Total	458	.02	.15	-.05	.16
i. Teacher Absence Proportion	Primary	244	.06	.06	-.17	.12
	Inter.	214	.06	.06	-.09	.05
	Total	458	.06	.06	-.12	.08
j. Proportion Free Lunch	Primary	244	.60	.32	-.36	.27
	Inter.	214	.59	.33	-.27	.36
	Total	458	.60	.32	-.29	.31
k. Amount Reading on Own	Primary	244	1.71	.28	-.34	.85
	Inter.	214	1.75	.29	-.16	.82
	Total	458	1.73	.28	-.26	.83
l. Average Interest in Reading	Primary	244	1.94	.43	-.33	---
	Inter.	214	1.99	.43	-.22	---
	Total	458	1.97	.43	-.28	---
m. Ave. Need for Extra Help Rdg.	Primary	244	2.10	.42	.42	-.40
	Inter.	214	2.23	.42	.18	-.37
	Total	458	2.16	.42	.23	-.37
n. % Receiving Help Elsewhere	Primary	244	37.90	29.90	.22	-.09
	Inter.	214	46.60	30.60	-.01	-.02
	Total	458	42.00	30.30	.05	-.05

TABLE I CONTINUED

Name of Variable	Level*	# of Classes	Mean	SD	Correlation With: Posttest**	Interest
o. Learning Problems-Reading	Primary	244	2.2	1.31	-.11	.17
	Inter.	214	1.85	1.20	-.15	.10
	Total	458	1.99	1.27	-.08	.13
p. Learning Barriers-Behavior	Primary	244	.85	.92	-.06	.14
	Inter.	214	.92	.94	-.10	.13
	Total	458	.89	.93	-.09	.14
q. Learning Barriers-Medical Physical Health (Vision, Sickness, etc.)	Primary	244	.44	.73	.00	-.02
	Inter.	214	.35	.66	-.07	.11
	Total	458	.40	.70	-.00	.03
r. Average Level of 1st Rdg Material	Primary	244	5.32	1.89	.04	-.23
	Inter.	214	9.74	1.23	.10	-.29
	Total	453	7.39	2.74	-.27	-.10
s. SD of Level of 1st Rdg Material	Primary	244	1.44	.60	-.17	.14
	Inter.	214	1.45	.66	-.19	.29
	Total	458	1.44	.63	-.17	.21
t. Presence of Math Specialist	Primary	244	1.00	.00	99.00	99.00
	Inter.	214	.99	.10	-.06	.04
	Total	458	1.00	.07	-.01	.02
u. Teaching Situation-Proportion Open Space	Primary	244	.23	.42	-.13	-.04
	Inter.	214	.24	.43	-.03	-.17
	Total	458	.23	.42	-.09	-.10
v. Teaching Situation-Proportion Team Tchg	Primary	244	.10	.10	.06	-.07
	Inter.	214	.14	.35	.16	-.23
	Total	458	.12	.33	-.07	-.15

* * * * * *

TABLE I CONTINUED

Name of Variable	Level*	# of Classes	Mean	SD	Correlation With: Posttest**	Interest
The following variables were included in some regression analyses. Note the change in sample size.						
aa. Mean Class Absences	Total	374	11.45	3.76	-.10	.18
bb. Representativeness of Reading Program (%)	Total	374	99.45	3.07	.05	-.15
cc. Mean Duration of Students in Rdg Class (Ave. Scale Value)	Total	374	7.35	.42	.11	-.05
dd. Class Average Need for Extra Help in Reading	Total	374	2.16	.31	.23	-.36
ee. Presence of Aide	Total	374	.08	.27	.10	-.18
ff. Presence of Reading Specialist	Total	374	1.00	.05	-.03	-.01
gg. Teaching Situation Departmentalized	Total	374	.04	.19	-.09	.04
hh. Teaching Situation Multi-Age Class	Total	374	.14	.34	-.09	-.11
ii. Teaching Situation Multi-Level Class	Total	374	.22	.42	-.08	-.12
jj. Taught by Other Teacher	Total	374	.37	.48	-.03	.03
kk. Learning Problems-Math	Total	374	1.45	1.46	.02	.06
ll. Learning Barriers-Economic Conditions (Clothing, Shoes, etc.)	Total	374	.28	.54	.03	-.01

a. Pretest % Objectives Mastered. This is the average of the class percentages of objectives mastered on the Prescriptive Reading Test (Fall 1977). The number of objectives mastered at a given test level was converted to the percentage of total objectives possible for that level. This was done for each student with matched pre-post test data for the same test level. The percentages for all matched levels in a given class were combined to give a weighted average for the class. The mean percentages in Table 1 are the unweighted average of class averages for the designated groups. The correlation shown for this variable with posttest is the correlation of the arcsine transformation of pretest with the arcsine transformation of posttest. A class is a homeroom class, since that is how students are grouped for testing.

b. Posttest % Objectives Mastered. Defined same as Pretest % Objectives Mastered, except pertains to posttest (Spring 1978). This is one of two main outcome measures used in the studies reported here.

c. Grade. This variable is the grade level of the class. The means in Table 1 are the average grade levels of the groups designated. The grade level of multi-grade classes was obtained by taking the unweighted average of the grades included in a given class. A weighted average would have been obtained by taking into consideration the number of students at each grade level within the class.

d. Test Level Count. This is the number of levels of PRT administered pre and post(matched) in a class. The means in Table 1 are the unweighted averages for the designated groups. The measure for a given class does not take into account the number of students receiving each test level.

e. Class Size. This is the total number of students who were recorded in the Evaluation System as members of a given class according

to certain inclusion rules. A student was considered a member of a class if the data for the student could be matched with a valid teacher and school ID number, and if the student had pretest and posttest data at the same test level. A class was that group of students having those characteristics plus an average membership duration of 100 or more school days in 1977-78, based on the collectivity of students with characteristics just defined. Thus, "class size", as the term is used here, is not the actual class size from the teacher's point of view, but the collectivity of students meeting certain criteria for purposes of analysis. In the sample of 458 classes for which data are presented in Table 1, most classes had sizes (as we have defined size) of 25 students. The reader needs to keep in mind that the term "class size" has a special meaning in this report.

f. % Males. This is the number of males divided by the number of members of a class (as defined in Class Size) times 100. It is the complement of the percentage of females.

g. Class Holding Rate. This is the complement of the class turnover rate. The turnover rate was obtained essentially by dividing the number of students who were not members of a class for the whole school year by the total number of members, and multiplying by 100. That number was then subtracted from 100.

h. Teacher Turnover Proportion. Any class that had a teacher who did not remain with a class for the school year was coded 1; all others were coded 0. The means in Table 1 are the proportions of classes in the designated groups that had a turnover of teachers by that definition.

i. Teacher Absence Rate. This is the ratio of the number of days the teacher was absent to the total number of school days counted for the year for the teacher. For 1977-78 the Evaluation System terminated

data collection on school day #162; that is the maximum number of teacher days possible for purposes of computing this measure.

j. Proportion Free Lunch. This is the number of students in a class reported by a teacher to be eligible for the free lunch program divided by the aggregate number of students constituting a class as defined in Class Size. This measure was intended to provide an index of the socio-economic composition of each class. It is important to recognize that the basic datum in this index is the number designated by the teacher for the whole class at a given point in time, not a designation for each individual student.

k. Amount of Reading on Own. This is the average for the class of the teacher's ratings of each student on the following scale:

"Aside from reading instruction, how often does this student read on his own in school? (library books, comics, etc.)"

1 = Often

2 = Sometimes

3 = Never

Ratings were made in April 1978.

l. Average Interest in Reading. This is the average rating for the class made of each student by the teacher on the following scale:

"How would you assess this child's interest?"

1 = Very interested

2 = Moderately interested

3 = Seldom interested

4 = Not interested

5 = Actively dislikes reading

This measure, like other ratings of individual students by their teachers, was made in the late Spring of 1978. Note in Table I that the

scale values, as with those for the preceding measure, are such that the more positive the rating, the lower the scale value. Thus the correlations of this measure with Posttest % Objectives are consistently negative; the correlations of this variable with Amount of Reading on Own are consistently positive (and high). One needs to bear in mind the direction of scales when one looks at the correlations in Table I in order to state the relationships correctly. In Chapter V, correlations involving these measures have had their signs reversed to aid the reader in making more direct interpretations.

This is the second main outcome measure used in the analyses reported here.

m. Average Need for Extra Help in Reading. This is the average, for the class, of ratings made of each student by the teacher on the following scale:

"Is this child in need of extra help in reading?"

1 = Yes, very much

2 = Yes, somewhat

3 = No

As with all such measures, the means, standard deviations and correlations of this measure are based on the distributions of it in the group designated in Table 1.

n. % Receiving Help Elsewhere. This is the number of students in a class rated by the teacher as receiving extra reading instruction during the regular school day from one or more sources other than the teacher, divided by the number of students rated; the ratio was multiplied by 100 to make it a percentage.

o. Learning Problems-Reading. This is the number of students in a class rated by the teacher as having been referred for reading difficulties on the following scale:

1 = 1-3 Students

2 = 4-6 Students

3 = 7-9 Students

4 = 10 or more Students

Note that the ratings for the measure are strictly scale values; the measure does not come from ratings of individual students.

p. Learning Barriers-Behavior. This measure is based on the same scale and procedure as Learning Problems-Reading.

q. Learning Barriers-Medical/Physical Health. This measure is based on the same scale and procedure as Learning Problems-Reading.

r. Average Level of First Reading Material. This is a rating made for each student by the teacher of the level of the reading instruction material used most frequently with the student during the month preceding the rating, which occurred in the late spring of 1978. Level was based on the following scale:

1 = Reading readiness

2 = Pre-primer

3 = Primer

4 = 1^1 (grade one, first semester)

5 = 1^2 (grade one, second semester)

6 = 2^1

7 = 2^2

8 = 3^1

9 = 3^2

10 = 4

11 = 5

12 = 6

13 = 7

14 = 8

15 = 9

s. SD of Level of First Reading Material. This is the standard deviation of the levels of the most frequently used reading instruction materials for the class. It was included in analyses as a measure of the heterogeneity of reading instruction for each class. The means and standard deviations of this measure in Table I pertain to the distributions of class standard deviations in the designated groups of classes.

t. Presence of Math Specialist. The school indicated whether there was a math specialist available to the class by checking Yes =1 or No =2 .

u. Teaching Situation-Proportion Open Space. The school indicated whether or not the class was open-space (Yes = 1, No = 0). The means in Table I are the proportions of open-space classes in the groups designated.

v. Teaching Situation-Proportion Team Teaching. Same code values and procedure as for Teaching Situation-Proportion Open Space. These variables are not mutually exclusive. The teacher could indicate the presence or absence of up to 5 teaching situation variables.

With the exception of variable aa below, the remaining variables were included in exploratory regression analyses, but were not included in the set of variables used in the regression analyses based on classes as the units in Chapter V of this report. They are, nevertheless, descriptive of classes in a total useable sample of similar size to that finally used. Thus, they are included in the final part of Table I.

aa. Mean Class Absences. This is the average number of days that students in a class were reported to have been absent during the 1977-78 school year for the Evaluation System schools. The average pertains to a maximum of the first 162 school days of that year. It was obtained by dividing the total number of days absent of all students in the class by the total number of days all students were members of the class (up to a limit of 162 days.

. SD of Test Level. The standard deviation of test levels for each class. Levels of the PRT were numbered (A = 1, B = 2, C = 3 ...) and treated as if they represented points on a continuum. The standard deviation for each class was investigated as a measure of class diversity of reading levels. (Not included in this analysis.)

bb. Representativeness of Reading Program. Teachers were asked to rate each student's reading program on the following scale:

> "To what extent is the reading program described above indicative of what was used all year for this child for reading instruction ? (Exclude changes in reading level.)"
>
> 1 = Don't know-student new to this class
>
> 2 = Very representative or very similar
>
> 3 = Somewhat representative or somewhat similar
>
> 4 = Not at all representative or not at all similar

A measure of representativeness for each class was obtained by dividing the total number of 2's and 3's by the total number of 2's, 3's and 4's and multiplying by 100. As can be seen in Table 1, the average of these class measures for 374 classes, grades 1-6, was very high (99.45%).

cc. Mean Duration of Students in Reading Class. (Average Scale Value.) On the Reading Program Form (DCDRE #200) for each student the

teacher rated how long the student had been in the class that year on the following scale:

1 = less than one month

2 = 1 month

3 = 2 months

4 = 3 months

5 = 4 months

6 = 5 months

7 = 6 months

8 = 7 months

9 = 8 months

The measure for each class is the average of ratings of time in class for the class.

dd. Class Average Need for Extra Help in Reading. This is the average for the class of the rating for each student on the following scale:

"Is this child in need of extra help in reading?"

1 = Yes, very much

2 = Yes, somewhat

3 = No

ee. Presence of Aide. Each school indicated on a form whether or not the class had an aide.

ff. Presence of a Reading Specialist. Each school indicated whether or not a reading specialist was available to the class (1 = Yes, 2 = No).

gg. Teaching Situation-Departmentalized. Same procedure as for variables u and v.

hh. Teaching Situation - Multi-Age Class. Same procedure as for variables u and v.

ii. Teaching Situation - Multi-Level Class. Same procedure as for variables u and v.

jj. Taught by other Teacher. Teachers indicated the number of their homeroom students who were given regular reading instruction by some other teacher .

kk. Learning Problems - Math. Same procedure as for variable o.

ll. Learning Barriers - Economic conditions. (Clothing, shoes, etc..) Same procedure as for variables p and q.

2. Benchmark Characteristics of Classes by Grade Level.

In order to provide benchmarks against which to compare previous reports, selected statistics were computed for classes in each of the six basic grade levels in the Evaluation System. These benchmark statistics are shown in Table II.

Of the 548 classes in the total useable sample 489 were used in Table II. Only single grade level classes were included. That is, multi-level classes were not included. The variables in Table II are the same as those defined above for Table I. Therefore, no further description of these variables is needed for Table II. It is, perhaps, appropriate to remind the reader that classroom averages for pretest and posttest percentage of objectives mastered are composites of the percentages for the different test levels administered pre and post. As Table II shows for the variable Test Level Count, there was typically more than one PRT test level used in the classes at each grade level. The average number of different test levels used for all of the 489 classes in Table II was 2.1. Further,

TABLE II

BENCHMARK CHARACTERISTICS OF CLASSES BY GRADE LEVEL

Characteristics	Grade Level 1	2	3	4	5	6	1-6
Number of Classes	81	92	87	78	74	77	489
Pretest % Objectives Mast'd	56.98	52.93	43.58	38.70	45.13	49.20	47.90
SD	9.21	10.90	8.96	10.08	9.23	12.97	
r_{xy}†	.44	.63*	.34*	.66*	.55*	.59*	
Posttest % Objectives Mast'd	82.51	79.73	77.96	73.07	71.33	70.87	76.15
SD	8.21	9.72	9.28	8.94	7.73	12.84	
Pre-Post Difference	25.53	26.80	34.38	34.37	26.20	21.67	28.25
Class Size	23.99	24.99	25.14	25.36	25.71	26.61	25.28
SD	3.81	3.84	3.35	4.26	5.29	2.61	
r_{xy}	-.04	.18	-.21	.01	.10	.05	
Mean Class Absences	13.34	10.84	10.42	11.51	11.11	11.53	11.44
SD	3.86	3.27	3.14	3.69	3.96	3.73	
r_{xy}	-.11	-.23*	-.17	-.27*	.10	-.13	
% Males	52.69	51.50	50.07	49.67	49.22	50.39	50.39
SD	9.54	9.07	8.20	8.25	11.07	7.20	
r_{xy}	.12	.12	-.12	-.14	.06	.05	
Class Holding Rate	87.95	88.80	89.27	90.70	89.90	92.10	89.73
SD	7.97	7.30	8.42	6.89	8.16	7.20	
r_{xy}	-.02	.22*	-.05	.12	.08	.17	
Test Level Count	1.19	1.75	2.25	2.58	2.99	2.12	2.01
SD	.42	.60	.82	.85	.96	.94	
r_{xy}	-.02	.05	-.15	.05	.10	.14	

* $p \leq .05$

† r_{xy} is the correlation of the variable with Posttest % Objectives Mastered (without the arcsine transformation of Posttest).

the pretest and posttest averages are based only on students who took the same level of PRT pre and post.

The benchmark characteristics for the 1977-78 pool of classes at each grade level were essentially the same as for 1976-77 and 1975-76 with one major exception. In 1977-78, the class average percent of reading objectives mastered on pretest was dramatically lower for grades 2-6 than in the two preceding years.

PRETEST AVERAGES:

	Grade						
Year	1	2	3	4	5	6	1-6
1977-78	56.98	52.93	43.58	38.79	45.13	49.20	47.90
1976-77	58.96	57.24	63.82	57.07	59.93	58.17	59.04
1975-76	55.94	60.92	66.65	62.15	60.38	59.68	61.00

POSTTEST AVERAGES:

	Grade						
Year	1	2	3	4	5	6	1-6
1977-78	82.51	79.73	77.96	73.04	71.33	70.87	76.15
1976-77	78.66	78.61	79.92	69.80	69.13	68.48	73.64
1975-76	81.17	80.87	79.13	74.77	70.28	65.12	75.00

Compared to previous years, apparent gains were markedly higher in 1977-78. Pre- and posttest variances were consistently smaller in 1977-78 than before, as were pre-post correlation coefficients (except at grade 2).

A two-way analysis of variance (ANOVA) was made using pretest scores as the dependent variable and year (1977-78, 1976-77) and grade as the independent variables. The results were highly significant:

Source	SS	df	MS	F	p
Within cells	111256.59	920	120.93		
Year	30007.18	1	30007.18	248.14	.001
Grade	8448.73	5	1689.75	13.97	.001
Year x Grade	10707.99	5	2141.60	17.71	.001

A one-way ANCOVA for each grade level (using year as the independent variable and pretest as the dependent) gave the following results:

Grade	F	df	p
1	.917	1,151	.340
2	9.750	1,163	.002
3	127.905	1,153	.001
4	128.582	1,154	.001
5	79.423	1,148	.001
6	18.032	1,151	.001

As noted, the marked decrease in pretest averages, with the general consistency or stability of posttest averages compared to prior years, radically affected average gains in percentages of objectives mastered (post minus pre):

	Grade						
Year	1	2	3	4	5	6	1-6
1977-78	25.53	26.80	34.38	34.37	26.20	21.67	28.25
1976-77	19.70	21.37	16.10	12.73	9.20	10.31	14.60
1975-76	25.23	19.95	12.48	12.62	9.90	5.44	14.00

A consequence of this change in 1977-78 was that, unlike prior years, the average gain for the intermediate grades was about the same as for the primary grades (27.46 compared to 28.94, respectively).

The decrease in average pretest percent of objectives mastered, and in pre-post correlations, is puzzling. There is no immediate information to account for it. The fall testing in 1977 was a month earlier than in 1976, but only a week earlier than in 1975. Other benchmark variables pertaining to the classes in the Evaluation System (see Table II) were not different from prior years. Thus, there seems no reason to believe there was a marked change in sample characteristics, nor an inadvertent change in aggregating procedures.

Questions must also be raised about the meaning of fall to spring gains. As long as pretest levels have remained relatively steady, the gains have seemed to be interpretable indicators of system performance. The abrupt drop in 1977-78 pretest levels, however, makes cross-year comparisons difficult, especially when the drop is accompanied by no particular change in posttest levels on a grade by grade basis. Perhaps it was that latter lack of change that was most perplexing of all.

It will be particularly important to see the pre-post results for 1978-79 and for 1979-80 to clarify expectations about performance levels. It will be necessary to bear in mind, however, that changes in the PRT design were introduced in 1978-79. Thus, direct comparisons with the 1977-78 data will not be possible, only indications of relative levels of performance pre and post.

D. Characteristics of Reading Programs

Since Chapter VI presents comparisons of the effectiveness of selected reading programs, it is appropriate to provide descriptions of the relative prevalence of programs. For purposes of this report, a reading program is defined as the types of reading instruction materials used most

frequently, with each student for the month preceding the completion of the Reading Program Form by the student's teacher near the end of the spring semester.

The Reading Program Form (see Appendix B) lists reading instruction material under seven major categories:

- Basal Series
- Linguistic/Phonic Materials
- Programmed/Structured Materials
- Language Experience Materials
- Individualized Reading Materials
- Supplementary Materials
- Special Materials

In describing a student's reading program, the teacher indicated the materials used most frequently, next most frequently, and third most frequently. For each material reported, the teacher also gave the reading level of the material (see definitions of variables beginning on page 16). Each material listed on the Reading Program Form falls uniquely into one of the seven main types or categories of materials listed above. Thus, one way of identifying a reading program is to consider only the main categories of materials listed for each student. Only first and second most frequently used material categories were used to define programs in reports of the 1975-76, the 1976-77, and the 1977-78 data.

Table III shows the prevalence of different programs, as difined, for grades 1-6. Prevalence here means the percent of students at a grade level classified as having had a particular program based on the main categories shown above. Table III also shows the programs selected for comparison at each grade level (see Chapter VI).

TABLE III

PERCENTAGES OF STUDENTS IN DIFFERENT READING PROGRAMS, BY GRADE AND YEAR 1/

Program	Year	Grade 1	2	3	4	5	6	1-6
Basal	'77-8	4.9	8.1	9.1*	8.3*	6.1*	7.1*	7.3
	'76-7	8.8*	7.7*	12.6*	10.9*	10.8*	7.5*	9.6
	'75-6	12.7*	10.1*	10.9*	7.4*	15.3*	12.8*	11.7
Basal-Ling/Phon	'77-8	31.2*	29.8*	31.4*	21.5*	32.2*	23.4*	28.3
	'76-7	19.0*	21.3*	22.9*	21.8*	24.9*	24.0*	22.3
	'75-6	35.4*	31.4*	30.5*	33.9*	27.0*	28.8*	30.8
Basal-Programmed	'77-8	4.2*	3.1*	5.0*	0.7	0.8	0.2	1.5
	'76-7	4.3	4.0	21.0*	0.7	1.7	0.5	5.0
	'75-6	8.5*	7.2*	10.0*	2.9	0.5	1.5	4.7
Basal-Lang. Exp.	'77-8	8.6*	6.2*	2.3*	2.9	3.8	0.0	4.1
	'76-7	9.0*	6.9*	4.9*	2.2	3.2	0.8	4.5
	'75-6	9.2*	6.6	3.2	4.2	2.4	0.2	4.0
Basal-Ind. Rdg	'77-78	2.8*	7.2*	10.2	20.3	11.4	10.1*	10.3
	'76-7	0.7	8.7	9.3	13.1	10.0	7.1*	8.1
	'75-6	0.8	7.0	14.8	17.3	22.6	8.6*	11.7
Basal-Supplementary	'77-8	5.9*	3.7	12.6	11.[illegible]	10.2	17.7*	11.0
	'76-7	2.9	3.8	4.[illegible]	11.1	6.4	11.8*	6.9
	'75-6	7.9	9.4	1[illegible].1	21.3	13.0	15.9*	13.2

TABLE III CONTINUED

Program†	Year	Grade 1	2	3	4	5	6	1-6
Basal-Special	'77-8	2.0*	0.0	---	1.0	0.0	0.5	
	'76-7	0.3	0.1	1.2	0.2	0.0	---	
	'75-6	---	---	---	---	---	0.1	
Ling/Phon	'77-8	7.7*	6.4*	3.1*	4.0	4.3*	1.4	4.5
	'76-7	6.5*	4.8*	5.4*	4.3*	2.3*	4.8*	4.7
	'75-6	3.2	7.8*	3.2	6.1	1.7	6.6*	4.6
Ling/Phon-Basal	'77-8	1.6	2.0*	0.3	0.6	0.0	1.3	
	'76-7	1.1	3.1	1.2	0.2	1.0	0.7	
	'75-6	6.5	2.7	1.8	0.3	0.7	0.8	
Ling/Phon-Prog'd	'77-8	---	---	---		---	2.0	
	'76-7	0.9	0.5	---		0.3	0.6	
	'75-6	3.0	1.5	0.2		0.1	0.7	
Ling/Phon-Lang. Exp.	'77-8	0.5	0.2	0.1	1.2	0.1	---	
	'76-7	0.8	0.1	0.0	0.4	1.3	0.0	
	'75-6	0.1	---	---	---	0.6	---	
Ling/Phon-Ind. Rdg.	'77-8	0.0	0.8	2.6	1.2	1.0	---	
	'76-7	1.7	0.8	1.6	3.0	2.2	0.7	
	'75-6	---	1.2	3.6	4.2	3.2	2.5	
Ling/Phon-Suppl.	'77-8	1.4	0.9	[illegible]	[illegible]	4.7	2.0	
	'76-7	2.3	3.2	[illegible]	[illegible]	3.1	2.6	
	'75-6	0.9	0.3	[illegible]		1.2	0.1	

TABLE III CONTINUED

Program	Year	1	2	3	4	5	6	1-6
Ling/Phon-Special	'77-8	---	---	---	---	0.1	0.3	
	'76-7	0.5	0.0	0.1	0.4	0.2	0.3	
	'75-6	---	---	---	---	0.1	---	
Programmed/Struct'd	'77-8	---	0.7	0.8	0.0	---	---	
	'76-7	0.6	1.4	0.5	0.2	0.2	0.1	
	'75-6	---	---	0.1	---	---	0.1	
Prog'd-Basal	'77-8	0.6	---	---	0.1			
	'76-7	0.2	0.2	2.1	---			
	'75-6	---	---	0.1	---			
Prog'd Ling/Phon	'77-8	---	0.7	---	0.1	1.3	2.8	
	'76-7	0.2	1.2	0.1	---	0.1	0.4	
	'75-6	---	---	1.5	---	---	---	
Prog'd Lang. Exp.	'77-8	0.3	---	0.8		---	---	
	'76-7	0.6	0.1	0.0		0.1	---	
	'75-6	0.1	---	---	---	0.1	0.2	
Prog'd-Ind. Rdg	'77-8	0.1	---	---	---	---	---	
	'76-7	---	0.1	0.0	0.0	0.2	0.1	
	'75-6		---	---	0.4	0.1	1.0	
Prog'd-Suppl	'77-8	1.0	---		0.1	0.7	2.0	
	'76-7	0.3	0.4		0.3	0.1	0.8	
	'75-6	---	---		---	---	0.3	

TABLE III CONTINUED

Program	Year	Grade 1	2	3	4	5	6	1-6
Prog'd-Special	'77-8	---	---	---	---	0.7	0.6	
	'76-7	0.1	0.0	0.2	0.3	---	0.1	
	'75-6	---	---	---	---	---	---	
Language Experience	'77-8	0.1	0.1	---	0.2		---	
	'76-7	0.4	0.1	---	---		0.1	
	'75-6	---	0.1	---	---		---	
Lang Exp-Basal	'77-8	0.2	0.1	0.0	0.8			
	'76-7	0.6	---	---	---			
	'75-6	0.2	0.2	0.6	0.1			
Lang Exp-Ling/Phon	'77-8	0.9	---	0.1	0.0			
	'76-7	0.0	---	---	---			
	'75-6	0.5	0.4	0.2	---			
Lang Exp-Prog'd	'77-8	---			0.0			
	'76-7	0.0			---			
	'75-6	---			---			
Lang Exp-Ind. Rdg	'77-8	0.0	---	0.0	---	0.0	---	
	'76-7	0.3	0.1	---	0.1	0.0	0.1	
	'75-6	---	---	---	---	0.1	0.2	
Lang Exp-Suppl.	'77-8	0.3	0.6	---		0.1		
	'76-7	0.3	---	0.7		---		
	'75-6	0.4	---	---		---		

TABLE III CONTINUED

Program†	Year	Grade 1	2	3	4	5	6	1-6
Indiv'd Reading	'77-8	---		0.0	0.1	0.0	---	
	'76-7	---		0.0	0.1	1.0	2.4	
	'75-6	---		---	0.1	---	0.2	
Ind. Rdg-Basal	'77-8	---	0.2	0.1	1.2	0.9	---	
	'76-7	0.1	0.1	1.3	1.1	1.0	0.0	
	'75-6	---	0.0	---	---	---	0.1	
Ind. Rdg-Ling/Phon	'77-8		0.8	1.0	1.0	---	---	
	'76-7		---	0.4	---	1.0	1.6	
	'75-6		---	---	0.1	2.6	0.1	
Ind. Rdg-Prog'd	'77-8	---					0.0	
	'76-7	0.0					---	
	'75-6	---					---	
Ind. Rdg-Lang. Exp.	'77-8					---	0.5	
	'76-7					1.4	0.6	
	'75-6					1.9	1.3	
Ind. Rdg-Suppl.	'77-8		---	---	0.0	1.2	1.5	
	'76-7		0.1	0.1	0.7	1.7	0.6	
	'75-6		---	---	---	0.1	1.8	
Ind. Rdg-Special	'77-8						---	
	'76-7						---	
	'75-6						0.1	

TABLE III CONTINUED

Program	Year	Grade 1	2	3	4	5	6	1-6
Supplementary	'77-8	0.0	---	---	---	---	0.6	
	'76-7	0.7	1.2	0.0	---	0.1	3.2*	
	'75-6	0.1	---	0.2	0.1	1.2	1.2	
Suppl-Basal	'77-8	0.2	---	---		1.9	0.0	
	'76-7	---	0.1	1.1		---	0.3	
	'75-6	0.2	2.2	---		0.3	---	
Suppl-Ling/Phon	'77-8	---			0.2	0.3	0.1	
	'76-7	0.1			0.0	0.0	2.2	
	'75-6	---	0.5		0.1	0.5	3.4	
Suppl-Prog'd	'77-8						0.1	
	'76-7						---	
	'75-6						---	
Suppl-Lang. Exp.	'77-8	0.3			0.0	0.0	0.0	
	'76-7	0.0			---	0.1	0.1	
	'75-6	0.1		---		---	0.1	
Suppl-Ind. Rdg.	'77-8	---		0.0		0.3	0.0	
	'76-7	---				1.1	0.5	
	'75-6	---				1.7	2.1	
Special	'77-8	1.1			0.1	0.0	0.9	
	'76-7	0.2			---	0.8	0.6	
	'75-6	---				0.1	---	

TABLE III CONTINUED

Program	Year	Grade 1	2	3	4	5	6	1-6
Special-Basal	'77-8					---	---	
	'76-7					0.0	1.0	
	'75-6					---	---	
Special-Ling/Phon	'77-8	---	---	---	0.6	0.2	0.2	
	'76-7	0.1	0.0	---	---	---		
	'75-6	---	---	0.1	---	---		
Special-Prog'd	'77-8							
	'76-7							
	'75-6							
Special-Ind. Rdg.	'77-8						---	
	'76-7						0.0	
	'75-6						---	
Special-Lang. Exp	'77-8		0.7			---		
	'76-7					---		
	'75-6					0.1		
Special-Suppl.	'77-8		0.0		0.0	0.2	---	
	'76-7						0.1	
	'75-6						---	
Unknown	'77-8	21.7	22.4	19.8	19.7	23.8	23.6	21.8
	'76-7	36.3	33.9	27.3	26.4	26.7	23.7	
	'75-6	10.1	11.0	8.0	7.6	3.6	9.5	

TABLE III CONTINUED

Program	Year	Grade 1	2	3	4	5	6	1-6
Total No. Students	'77-8	2396	3081	2719	2502	2452	2569	15719
	'76-7	2782	2736	2341	2724	2611	2859	16053
	'75-6	952	1104	1206	1204	1558	1227	7251

*Programs selected for comparison within a grade level (see Chapter VI).

Abbreviations: Ling/Phon = Linguistic/Phonics
Programmed or Prog'd = Programmed/Structured
Lang. Exp = Language Experience
Ind. Rdg = Individualized Reading
Suppl = Supplementary

1/ The percentages for 1975-76 and 1976-77 will appear different from those given in the two previous reports because they have been recomputed based on total count which includes the category "unknown".

In Table III the order in which categories are listed for a given program is: most frequently used, next most frequently used. Also, some programs appear to involve only one category of material; e.g. Basal, Linguistic/Phonic, and Language Experience. Such programs mean that the teachers indicated either just a single material, or two materials falling in the same category; for example, a basal text and nothing else; or two different basal texts. The percents are based on the total number of student files in each grade, including files with missing data.

It can be seen in Table III that the most prevalent type of program (in terms of percentages of students) is a combination of Basal and a Linguistic/Phonics materials. That is true for all six grades.

The prevalence of different types of reading programs for 1976-77 and 1977-78 was generally similar. There were some changes, however, that may be indicative of changing trends in the utilization of different theories and methods of reading instruction. In 1977-78, the use of Basals without other types of materials was down at all grade levels except grade 2. The use of a Basal along with a Linguistic/Phonic material increased in 1977-78, particularly in the primary grades. There was a decrease in the reported use of Programmed/Structured materials to supplement Basals, especially at the third grade level. Finally, compared to 1976-77, the use of a Basal-Supplementary materials type of program was more prevalent, especially in the intermediate grades.

These shifts in the prevalence of different types of reading materials used suggest important changes in the theories and methods of reading instruction occurring throughout the school system. The consequences of these changes will reveal themselves over time in the achievement of students. Analyses of the 1978-79 data will be made to determine whether these trends continue, and if they do, how they are affecting reading achievement.

V. RESULTS BASED ON REGRESSION ANALYSES

The earlier Evaluation System Reports on the 1975-76 and 1976-77 data contained a number of regression analyses. This section of the present report provides a parallel set of regressions on the 1977-78 data base. In what follows, we report first on the nature of regression analysis and the particular variables included in our analyses. This is followed by a section on the specific regression analyses carried out on the 1977-78 data. Results for the three years are treated comparatively at appropriate points.

The regression analyses were carried out in an effort to determine what variables, from within the set described above, are significantly associated with and predictive of reading performance and reading interest. The practical implications of such an approach to data analysis are straightforward. Basically, such analyses can help us to determine what prior test performance characteristics, classroom characteristics, teaching practices, and student characteristics actually "make a difference" in the sense of being statistically associated with reading performance and interest. Knowledge of which variables are predictive of performance and interest scores will provide for the school system an empirical basis for making possible modifications which can reasonably be expected to lead to increases in either reading performance or reading interest or both. It is important to keep in mind that the results reported in this chapter refer to classes, not to individual students.

A. The Regression Analyses

In order to accomplish our basic purpose of learning what sets of variables and what variables within those sets are significantly predictive

of reading performance and interest scores, a series of regression analyses were performed. In each regression analysis, there is a single variable, termed the criterion variable, which we are trying to predict and a number of different variables that we have reason to believe may be associated with the criterion. These latter variables are termed predictors and are examined, individually and in sets, to determine which ones do, in fact, significantly predict the criterion. To clarify the nature of the regression analysis, consider briefly the example of a baseball manager who wishes to increase the batting averages of his players. Taking a systematic regression approach, he might determine the existing batting averages of a large number of players and collect information concerning each player's height, weight, muscle volume, amount of experience, age, and the training methods employed in teaching that player to bat. He would enter all of these variables into a regression analysis to determine which would significantly predict the criterion "score," the player's batting average. With this approach, he can learn which variables are important in the training of players in the future and may thus learn, for example, that players with greater height, but not necessarily with greater weight, who have greater muscle volume and have been trained with certain training methods have higher batting averages.

1. The Criterion Variables

The "batting averages" of interest in the present study were the posttest reading performance score and the student's interest in reading as rated by the teacher. The reading performance score employed was basically the percentage of objectives mastered (Posttest ÷ Objectives). This percentage was determined for each individual student, and the

individual percentages were averaged to determine the average or mean score for each class. This percentage score, termed the posttest score, is described in more detail in Chapter IV.[5/]

The second criterion score employed in separate regression analyses was the student's interest in reading. This score is based on an observer rating scale, completed by the student's teacher as an item on one of the forms described in Chapter II. For each student, the teacher was asked to assess the child's interest in reading by indicating which of the following best described the student's interest level:

1 = Very interested

2 = Moderately interested

3 = Seldom interested

4 = Not interested

5 = Actively dislikes reading

For statistical purposes, each of these terms represented a point on a 5-point rating scale, where a high score of 5 is assigned to "very interested" and a low score of 1 to "actively dislikes reading" (intentionally reversed).

2. Predictor Variables

The variables used in an attempt to predict the posttest performance and reading scores have been described. For purposes of analysis, the variables finally selected for entry were grouped into nine major sets. In fact, four of these "sets" consisted of a single variable, while the remaining 5 contained between 2 and 5 variables each. Following is a brief description of each set and a listing of the variables within that

---

5/ The posttest score was subjected to an arcsine square root transformation before regression analyses were carried out.

set. Sets and variables are described in the order in which they were entered into the hierarchical regression analyses. For a more detailed treatment of the individual variables and the distributional characteristics associated with these variables, the reader is referred back to Chapter IV.

a. Set 1 - Pretest Score. A single-variable set, pretest score is the class average percentage of correct objectives and represents the fall reading performance test score of the individual child averaged over the classes as a whole.[6/] The entry of the pretest score at an early point in the hierarchical regression is important, since it served to partial out, and thereby control, the effect of the student's reading level at the beginning of the year when examining his/her reading level at the end of the year as measured by the posttest score.

b. Set 2 - Grade. This set consists of a single variable, and is simply the grade of which the class is a part. Grades 1 through 6 are represented.

c. Set 3 - Test Level Count. This set also consists of a single variable, which is simply a count of the number of different test levels represented by children in a given class. It is, in effect, a measure of the degree of homogeneity of grouping, where homogeneity is defined in terms of test levels. A very homogeneous class might have only a single level of testing represented in the entire class (all students at the same level), while a relatively heterogeneous class might have several different levels represented.

6/ As with the posttest score described above as a criterion variable, the pretest score was subjected to an arcsine square root transformation before regression analyses were carried out.

d. Set 4 - Basic Classroom Characteristics. This fourth set consists of three variables, each contributing to a basic description of the class being entered into the analysis. In utilizing these variables, it was considered possible that such characteristics as class size might contribute significantly to the prediction of the posttest reading score or the reading interest score or both. The three variables in the set were:

- Class size
- Percent male (sex distribution)
- Class holding rate

e. Set 5 - Teacher Behavior Characteristics. This set of variables describes two potentially important aspects of teacher behavior, characteristics which we had reason to believe might be predictive of students' posttest scores. The two major variables included in the set were:

- Teacher turnover rate
- Teacher absence rate

f. Set 6 - Percent Free Lunch. This set, consisting of a single variable, indicates the percentage of students in a given class receiving free lunch. The variable may provide some indication of the socioeconomic level represented in the class and again may be predictive of reading interest or performance.

g. Set 7 - Reading Behavior Characteristics. Consisting of 4 major variables, the reading behavior set deals with certain student characteristics (averaged over the class as a whole) that are directly involved with reading and reading activities and may well be predictive of reading performance. The four variables are:

- Average amount of reading on own
- Average interest in reading
- Average amount of need for extra help in reading
- Average percentage receiving help elsewhere

h. Set 8 - Learning Problems and Barriers. Certain kinds of problems faced by the individual student may serve as barriers to effective learning. The inclusion of Set 8 in the analyses represents an attempt to examine the relationship of the prevalence of problems and barriers in a class to class performance or interest scores. The three major variables contained in the set are:

- Health barriers
- Reading problems
- Behaviors barriers

i. Set 9 - Classroom Environment. The learning experiences of the individual student in the classroom setting are, to some degree, determined by certain kinds of variables which, together, comprise what may be termed the classroom environment. The present analyses entered several of these potentially important variables in an attempt to identify those aspects in the classroom environment which may, in fact, relate significantly to reading performance and interest. The five major variables involved in this final set are:

- Mean level of most-used reading material
- Presence of a math specialist
- Open space versus self-contained classroom
- Team teaching vs. non-team teaching
- Variability (standard deviation) of level of most-used reading material

3. Specific Regression Analyses

As with the 1975-76 and 1976-77 data, a series of six regression analyses were carried out. The first three of these involved the class average posttest score as the criterion variable and the nine sets of variables noted above as predictors. The first analysis involved a valid sample of 458 classes (excluded classes missing data on specific variables) cutting across all six grades. It should be noted that this represents a substantial gain over the 229 classes available for analysis in the 1975-76 and the 396 in 1976-77. The larger sample size in the more recent data base reflects primarily continuing progress in the development and improvement of the Evaluation System. While it is important to know which variables and sets of variables significantly predict posttest across all grades, it was considered likely that somewhat different variables might significantly predict the posttest scores at primary and intermediate grades. As a result, a second regression analysis employed the posttest scores and associated predictors scores in the first three grades (1, 2 and 3), while a third analysis examined relationships in the three intermediate grades (4, 5 and 6).

The remaining three analyses paralleled those described above, except that the criterion variable was the class average interest in reading rather than posttest reading performance score. In addition, for these analyses, it was necessary to remove reading interest and average amount of reading, which closely parallels the interest score, as predictors. Average class absence rate was added as a variable.

4. Limitations

Before presenting the results of the regression analyses, it is important to recall that the EDACA system and the Evaluation System

generally are still being developed and tested. While there have been definite improvements in both the EDACA System and the data base for the current analyses, the data base still necessarily has certain limitations. The sample of 458 total classes considered valid for analysis is certainly larger than the samples of 229 and 396 classes available in the 1975-76 and 1976-77 data, respectively. However, there is still considerable data loss, and the loss may be systematic, rather than totally random. The problem is not a serious one, since we will be crossvalidating present findings in future years.

The results of each of the regression analyses on posttest performance scores, as well as those on reading interest scores, are provided in two separate sections. The first provides a brief statement of the statistical results of the analyses, while the second provides a discussion and interpretation of the results.

B. Regression on Posttest Performance Scores

As has been noted, three regression analyses involving the posttest reading scores as the criterion variable were conducted. The first cuts across all six grades, the second deals with the primary grades only, and the third deals with the intermediate grades only.

1. Means and Standard Deviations

Table I in Chapter IV provides the means and standard deviations for all variables across all grades and separately for the primary three and the intermediate three grades.

2. The All-Grades Regression

Results of the first regression analysis are dealt with in two separate sections, as noted above.

a. Results. Table A (in Appendix D) provides the matrix of correlations among the predictors and the criterion variable involved in the regression. Table IV indicates the multiple correlation (R), the R^2 value, the R^2 change, and the simple correlation associated with each variable in the fixed order regression. The final R and R^2 values indicate that the variables entered resulted in a significant correlation of .69, which accounts for approximately 48% of the variance. A more conservative estimate suggests a likely population R^2 of .47.

In order to provide an appropriately conservative estimate of variables contributing significantly to the regression equation, F-ratios were first calculated to determine whether or not a given set of variables made a significant contribution. Only if the set contribution was significant were the F-ratios for individual variables calculated. It should also be noted that, for both sets and individual variables, the F-ratios calculated were based on a Model II (Cohen & Cohen, 1975) approach, in which the error term involves all variables entering into the total analysis.

Table V provides the F-ratios and associated R^2 change values for the nine sets of variables entered into the regression. It is clear from the table that five of the nine sets significantly contributed to the prediction of posttest reading scores. The first set, Reading Pretest Score, was entered primarily as a control variable in order to partial out the effect of pretest performance on the further prediction of posttest scores. As would be expected, pretest is positively associated

Table IV

Posttest Regression Summary: All Grades

Variable	Multiple R	R Square	RSQ Change	Simple R	F-Ratio		B	Beta
Pretest Score	.56333	.31734	.31734	.56333	266.247***	†δ	.44969	.43570
Grade	.61574	.37914	.06180	-.38910	51.850***	†δ	-.03171	-.42641
Test Level Count	.61577	.37917	.00003	-.20782	SNS	†	.00360	.02773
Class Size	.61604	.37950	.00038	-.05142	SNS		.00030	.00891
% Males	.61707	.38078	.00128	.09166	SNS		.00074	.05313
Class Holding Rate	.61751	.38131	.00053	-.00644	SNS		-.00034	-.02071
Teacher Turnover	.62138	.38612	.00480	-.05062	4.027*		-.02773	-.03216
Teacher Absence	.63173	.39908	.01297	-.11753	10.882***		-.14739	-.07118
% Free Lunch	.65997	.43557	.03648	-.28753	30.615***	†	-.04195	-.10768
Reading on Own	.66986	.44872	.01315	-26156	11.033***	†δ	.00480	.01078
Reading Interest	.67320	.45319	.00447	-.28235	3.750	†δ	.03471	.11802
Need for Help	.67945	.46165	.00846	.23353	7.098**		.02833	.09501
% Receive Help	.68052	.46310	.00145	.04737	1.217		.01752	.04224
Health Barriers	.68344	.46709	.00399	-.00198	SNS		.00095	.04977
Reading Problems	.68354	.46723	.00014	-.07545	SNS		.00036	.00361
Behavior Barriers	.68385	.46765	.00042	-.08700	SNS		.00312	.02289
Math Specialist	.68577	.47029	.00013	-.00883	SNS	†	.01611	-.00843
Team Teaching	.69304	.48031	.00153	.07310	SNS		.01810	.04670
Open Space	.69194	.47878	.00849	-.08645	SNS	δ	-.03497	-.11745
Variability of Material Level	.69306	.48033	.00002	-.16619	SNS		-.00113	-.00562
Average Material Level	.68568	.47015	.00250	-.27026	SNS	†δ	.00643	.13940

Note: SNS means that the individual F was not calculated because the F for the set was nonsignificant.

*p < .05
**p < .01
***p < .001
† Significant in 1975-76
δ Significant in 1976-77

TABLE V

POSTTEST SCORE REGRESSION - ALL GRADES: CONTRIBUTION OF SETS

Set No.	Set Name	Number of Variables in set	F	1975-1976	1976-1977	R^2 Change
1	Reading Pretest	1	266.25***	+	δ	.317
2	Grade	1	51.85***	+	δ	.062
3	Number of Test Levels	1	0.03	+		.00003
4	Classroom Characteristics	3	0.60			.002
5	Teacher Characteristics	2	7.45***			.018
6	Percent Receiving Free Lunch	1	30.62***		δ	.036
7	Reading Behavior	4	5.77***	+	δ	.028
8	Learning Barriers	3	1.27			.005
9	Classroom Environment	5	2.13	+	δ	.013

* $p < .05$
** $p < .01$
*** $p < .001$

\+ significant in 1975-76 year
δ significant in 1976-77 year

with posttest and makes a reliable contribution to the regression equation. The second set, Grade, was also significant and, like pretest score, was entered primarily in order to partial out the effect of the variable.

The third significant set, Teacher Behavior Characteristics, contains two major variables. Each was individually significant. Teacher Turnover Rate and Teacher Absence Rate related negatively to posttest reading scores, indicating that classes with higher rates of turnover and absence have lower reading performance.

The fourth significant set was the percentage of students receiving free lunch. The statistical reliability of this rough measure of socio-economic level indicates that classes with higher proportions of low socioeconomic level students tend to have lower reading performance.

The final significant set was Reading Behavior Characteristics. All four individual variables in this set were significant. Classes with a high average amount of self-initiated Reading have higher posttest reading scores. Second, high average interest in reading in the class is predictive of high reading performance. Third, classes in which pupils have a high need for extra help in reading have lower reading performance scores. And, finally, classes in which a high percentage of students are receiving extra reading instruction outside the regular classroom instruction achieve higher average posttest scores.

b. Discussion and comparison. The regression on posttest scores across all grades showed that five major sets of variables are importantly involved in the prediction of these criterion scores. The first two, reading pretest score and grade, are control variables. The significance of the grade variable is based on the fact that there was a decrease in posttest scores with increasing grade. This result is consistent with those seen in both previous years (Table 5) and probably does not reflect an

actual decrease in reading abilities across grades, but rather a change in the characteristics of the test across its increasing levels. A fourth significant set of variables entering the equation, those dealing with teacher behavior characteristics, was not significant in previous years. Present data indicate, with regard to these variables, that classrooms in which there is a high rate of teacher turnover (i.e., the class has one or more changes of teacher per year) and classes in which the teacher is absent a higher proportion of the time are likely to have lower reading performance scores. This is an important finding, particularly with regard to teacher absence rates, since it has clear practical application. Specifically, the finding provides a statistical basis for discouraging teacher absences. The result should however, be interpreted with caution, since it was not significant in previous years. Before the significant roles of teacher turnover rates and teacher absence can be firmly established, it will be essential to replicate current results in data for future years.

The final significant set, Reading Behavior Characteristics, was also significant in both previous school years. Moreover, two of the specific variables in this set, average amount of self-initiated reading and reading interest, were also significant in all three years, indicating that these variables are particularly important in predicting reading performance. Clearly, the suggestion would be that teachers should encourage pupils to read more on their own and should do as much as possible to stimulate interest in reading in the classroom. It is recommended that methods of increasing reading interest and extra reading in the classroom be studied and that teachers be trained in the best available methods for accomplishing these goals, as the end result may be increases in reading performance.

The remaining two significant variables in the reading behavior set in 1977-78 data were not significant in previous years. For the current year, the need for extra help variable suggests that classes with a high need tend to be low in reading performance, as might be expected. More importantly, it was found that classes actually receiving extra help have higher reading posttest scores. This means that, at least for the 1977-78 data, the receipt of extra reading instruction does increase performance scores. Administratively, this may mean attempting to provide yet more resources for providing children with individualized reading instruction above and beyond that which can reasonably be provided by the classroom teacher. It is cautioned that administrative action on these last two variables may not be appropriate until they have been replicated in data for future years, since they were not significant in the first two years.

c. Multi-Year Consistency. One major reason for repeating the regression analysis with the same variables across a three-year period was to select for further study and action those variables showing the greatest degree of consistency across years. Some of the consistencies and inconsistencies have been noted above. One additional inconsistency to be noted is the fact that the classroom environment set, significant in both previous years, was not significant in the 1977-78 data. it may be that one or more of the five variables in that set would have been individually significant. However, under our conservative statistical approach, these individual variables were not examined in current data due to the nonsignificance of the set. It seems reasonable to suggest that the classroom environment set should be pursued further in future years, since it has been significant in two of the three years studied to date. A more general,

and perhaps more important, observation is that certain variables have been consistently significant in data for all three years. These include reading pretest score, grade level, amount of self-initiated reading, and reading interest. In addition, the percentage of students receiving free lunch (socioeconomic level) and the average level of reading material used in the classroom have been significant in two of the three years. It would seem reasonable to conclude that action relevant to variables consistent across two years is probably warranted.

3. Regression Analysis for Primary Grades (Grades 1-3)

Data for the first three grades were entered into a separate regression analysis similar to the one described above.

a. Results. Means and standard deviations for the predictor and criterion variables are provided in Table 1 (Chapter IV). Tables VI and VII provide the regression summary for separate variables and the F-ratios and associated R^2-change values for sets, respectively, for the regression analyses of the primary grades. Table B (in Appendix E) provides the correlation matrix. The final, significant multiple correlation for the primary grades was .72, accounting for approximately 51% of the variance. The estimated population R^2 based on the latter is .49.

Six sets of variables entered significantly into the final regression equation. They were, in order of entry into the hierarchical analysis, Pretest Score, Test Level Count, Teacher Behavior Characteristics, Percent Receiving Free Lunch, Reading Behavior Characteristics and Learning Barriers. The Pretest Set made a contribution, as a control variable, similar to that seen in the all-grades analysis above. Classes with high average pretest scores had higher average posttest scores.

The significant effect of test level count was due to the fact that classes with less test levels represented tended to have higher posttest

TABLE VI

POSTTEST REGRESSION SUMMARY: PRIMARY GRADES

Variable	Multiple R	R Square	RSQ Change	Simple R	F-Ratio		B	Beta
Pretest Score	.52681	.27753	.27753	.52681	126.628	†δ	.50308	.46853
Grade	.53102	.28199	.00445	-.18354	SNS	δ	.00726	.04849
Test Level Count	.54006	.29167	.00968	-.14995	4.417*		-.02095	-.13397
Class Size	.54144	.29315	.00148	-.00024	SNS		.00200	.06058
% Males	.54234	.29413	.00098	.07493	SNS		.00077	.05667
Class Holding Rate	.54274	.29457	.00043	.02844	SNS		-.00113	-.07374
Teacher Turnover	.54580	.29790	.00334	-.08105	1.524		-.02096	-.02658
Teacher Absence	.57258	.32785	.02995	-.17294	13.665***		-.23745	-.12399
% Free Lunch	.63242	.39995	.07210	-.36166	32.897***		-.04828	-.12648
Reading on Own	.65498	.42900	.02905	-.34046	13.255***	†δ	.01582	.03622
Reading Interest	.66117	.43715	.00815	-.33027	3.719	†	-.04019	-.14172
Need for Help	.68796	.47329	.03614	.42303	16.489***		.02709	.24802
% Receive Help	.69529	.48343	.01013	.21535	4.622*		.04269	.10258
Health Barriers	.70548	.49770	.01428	.00331	6.52*		.01818	.10895
Reading Problems	.70710	.49998	.00228	-.10529	1.04		.00462	.04944
Behavior Barriers	.70892	.50256	.00258	-.06371	1.18		.00563	.04226
Math Specialist	.70972	.50370	.00000	.00000	SNS		.00000	.00000
Team Teaching	.71181	.50668	.00055	.06233	SNS		-.01702	.04150
Open Space	.71142	.50612	.00242	-.13423	SNS		-.02056	-.07080
Variability of Material Level	.71655	.51345	.00677	-.17127	SNS		.01889	-.09261
Average Material Level	.70972	.50370	.00114	.03894	SNS		.00342	.05281

Note: SNS means that the individual F was not calculated because the F for the set was nonsignificant.

*p = .05
**p = .01
***p = .001

† Significant in 1975-76
δ Significant in 1976-77

TABLE VII

POSTTEST SCORE REGRESSION - PRIMARY GRADES: CONTRIBUTION OF SETS

Set No.	Set Name	Number of Variables in set	F	1975 1976	1976 1977	R^2 Change
1	Reading Pretest	1	126.63***	†	δ	.278
2	Grade	1	2.04		δ	.004
3	Number of Test Levels	1	4.42*			.010
4	Classroom Characteristics	3	.44			.003
5	Teacher Characteristics	2	7.59***			.033
6	Percent Receiving Free Lunch	1	32.90***			.072
7	Reading Behavior	4	9.52***	†	δ	.083
8	Learning Barriers	3	2.91*			.019
9	Classroom Environment	5	.99			.011

* $p < .05$
** $p < .01$
*** $p < .001$

† significant in 1975-76 year
δ significant in 1976-77 year

reading scores. It should be noted that this variable was not significant in the all-grades analysis.

The Teacher Behavior Characteristics Set contained one individually significant variable for the primary grades. This was teacher absence rate. Results showed that a high teacher absence rate is associated with relatively poor performance of the class as a whole on the reading posttest.

The final three significant sets were percent receiving Free Lunch, Reading Behavior Characteristics and Learning Barriers. The former, a single-variable set, showed that having a high percentage of students receiving free lunch in a given class is predictive of relatively poor reading performance in that class. Within the Reading Behavior Characteristics set, three variables were significant. First, classes with higher average amounts of self-initiated reading have higher reading performance. Secondly, those classes having a relatively small number of pupils needing extra help in reading have higher performance levels. Thirdly, classes receiving substantial amounts of outside help in reading attain higher performance levels by the end of the year. Only one variable in the significant learning barrier set reached significance. The result indicates that classes with more health problems tend to show relatively poor reading performance. This result has not been seen in previous analyses and should be regarded with caution. The final multiple correlation for the intermediate grades was .62, accounting for approximately 39% of the variance. The estimated population R^2 based on the latter is .36.

b. Discussion and comparison. The regression analysis for primary grades yielded six significant sets. The first set, pretest score, is a control variable. The second set, a single-variable set, was test level count. Here, results suggest that, in the primary grades, classes with large numbers of test levels represented have relatively low reading performance scores. Test level count can be taken as a rough measure of the homogeneity of grouping within classes. The direction of significance

of this variable suggests that more homogeneous grouping is associated with better performance. The variable should, however, be interpreted with caution, since it has not been consistently significant in analyses for previous years. It should be subjected to further study in future years before any action based on this result would be clearly warranted.

The teacher absence rate and percent free lunch variables, significant in the 1977-78 data, have not been significant for the primary grades in previous years. Both, however, are directionally consistent with expectations. The teacher absence variable, if replicated in data for future years, provides a statistical basis for encouraging teachers to minimize classroom absences. The percent free lunch variable has been considered, in previous reports, to be a rough indicator of socioeconomic level. For both the primary grades and all grades data in the present analyses, this variable is significant. Classes having higher percentages of free lunch students and therefore those lower in socioeconomic level tend to have lower reading performance scores. What action could be taken based on the significance of this variable is unclear. It would first be necessary to confirm, with other measures, the relevance of socioeconomic level to reading performance. Further study would then need to be undertaken to determine just what actions the school system could take that might overcome the negative influence of high proportions of low socioeconomic pupils in a given class.

The three significant variables in the Reading Behavior Characteristics set have considerably clearer practical implications. First, the self-initiated reading variable (amount of reading on own) is among the most consistently predictive variables across analyses done over the three-year research period. Larger amounts of self-initiated reading are consistently predictive of higher reading performance scores. The practical

implications of this finding have been considered and recommendations outlined in other reports. Briefly, it would be suggested that the school system take some action based on the significance of the self-initiated reading variable. Teachers might be informed of the importance of this factor and trained in classroom techniques that are likely to lead to an increase in extra reading. In addition, reading specialists and principals should be alerted to assure that there are substantial amounts of interesting extra reading materials available in each classroom, that pupils have an adequate amount of time in school to read these materials, and that they are encouraged to do so.

The remaining significant variables in the reading behavior set show that classes rated by teachers as having a high need for extra help tend to have reduced reading performance, as might be expected. However, it is of particular interest to note that those classes actually receiving larger amounts of extra reading instruction have higher posttest performance scores. Again, the practical implication is clear. If a class needs extra help, that help should be made avalable if at all possible. More generally, the availability of extra reading instruction is clearly desirable, since it is significantly associated with increased reading performance. Again, however, a word of caution is in order, since these two variables have not been consistently significant for the primary grades in previous years.

4. Intermediate-grade Regression Analyses (Grades 4-6)

The final regression analysis involving posttest reading scores was that for the three intermediate grades.

a. Results. Means and standard deviations for the intermediate grade variables are provided in Table 1 (Chapter IV). Tables VIII and IX provide, respectively, the regression summary of the F-ratios for sets

TABLE VIII

POSTTEST REGRESSION SUMMARY: INTERMEDIATE GRADES

Variable	Multiple R	R Square	RSQ Change	Simple R	F-Ratio		B	Beta
Pretest Score	.51370	.26389	.26389	.51370	82.776***	†δ	.51904	.57405
Grade	.57796	.33404	.07015	.07513	22.004***	†δ	-.03747	-.28049
Test Level Count	.57820	.33432	.00028	.02458	SNS	†	.00197	.01762
Class Size	.57822	.33434	.00002	.00932	SNS		-.00075	-.02496
% Males	.58093	.33747	.00313	.01783	SNS		.00080	.06377
Class Holding Rate	.58655	.34404	.00657	.08324	SNS		.00115	.07394
Teacher Turnover	.59145	.34981	.00577	-.03645	SNS		-.06487	-.07962
Teacher Absence	.59484	.35383	.00402	-.08620	SNS		-.06409	-.03327
% Free Lunch	.60197	.36237	.00854	-.27439	SNS		-.02642	-.07831
Reading on Own	.60418	.36539	.00303	.15935	SNS		-.00937	-.02421
Reading Interest	.60639	.36771	.00232	-.21788	SNS		-.02099	-.08066
Need for Help	.60775	.36936	.00165	.18248	SNS		-.02933	-.11019
% Receive Help	.60790	.36954	.00018	-.00992	SNS		-.01226	0.03425
Health Barriers	.60795	.36961	.00007	-.07315	SNS		-.00330	-.01974
Reading Problems	.61268	.37538	.00577	-.15090	SNS		-.00927	-.10073
Behavior Barriers	.61294	.37569	.00031	-.09925	SNS		.00143	.01210
Math Specialist	.61406	.37707	.00125	-.05924	SNS	†	-.04652	-.04057
Team Teaching	.62263	.38766	.00566	.15889	SNS		.03024	.09647
Open Space	.61806	.38200	.00493	-.03315	SNS		-.03073	-.11864
Variability of Material Level	.62281	.38790	.00023	-.18555	SNS		.00373	.02226
Average Material Level	.61304	.37582	.00013	.10382	SNS	†	.00221	.02449

Note: SNS means that the individual F was not calculated because the F for the set was nonsignificant.

*p < .05
**p < .01
***p < .001
† Significant in 1975-76
δ Significant in 1976-77

TABLE IX

POSTTEST SCORE REGRESSION - INTERMEDIATE GRADES: CONTRIBUTION OF SETS

Set No.	Set Name	Number of Variables in set	F	1975-1976	1976-1977	R^2 Changes
1	Reading Pretest	1	82.78***	†	δ	.264
2	Grade	1	22.00***	†	δ	.070
3	Number of Test Levels	1	.09	†		.0003
4	Classroom Characteristics	3	1.02			.010
5	Teacher Characteristics	2	1.54			.010
6	Percent Receiving Free Lunch	1	2.68			.009
7	Reading Behavior	4	.56			.007
8	Learning Barriers	3	.64			.006
9	Classroom Environment	5	.77	†		.031

*p < .05
**p < .01
***p < .001

† significant in 1975-76 year
δ significant in 1976-77 year

and statistics for individual variables. Table C (in Appendix F) provides the correlation matrix.

Results showed that only two sets were significant. These were the two control sets, pretest score, and grade. The pretest score variable, as usual, showed that high pretest scores are predictive of high posttest scores. The grade variable, as in some analyses for prior years, indicated a decrease in posttest performance scores with increasing grade. No other set was significant in the intermediate grade analysis. As a result, individual variables within sets were not further examined.

b. Discussion and Comparison. Results of the regression analysis for the intermediate grades were entirely consistent with those for the 1976-77 year, showing that only reading pretest and grade level are significantly predictive of posttest reading scores. Although two additional sets, test level count and classroom environment, had been significant in the 1975-76 year, the consistent results for the two later years suggest, overall, that the set of variables currently available on the Evaluation System is more useful in predicting reading performance for the primary grades than for intermediate levels. It may be necessary for the Evaluation System to identify additional variables with greater potential predictive power that can be assessed in future years for intermediate grade classrooms.

C. Regression Analysis of Reading Interest

As in data for previous years, a series of three hierarchical regression analyses, paralleling those done on posttest performance scores, were undertaken with regard to teacher ratings of interest in reading. Thus, for these analyses, the criterion variable was a reading interest score based on teacher ratings as described above. The all-grades, primary

grades and intermediate grades analyses are reported below.

1. Basic Statistics

Table I (in Chapter IV) provides the complete set of means and standard deviations for the criterion and predictor variables.

2. Analysis of All Grades

As in the first posttest regression analysis reported above, data from all six grades were entered into the overall regression analysis. All valid classes were included.

a. Results. Table A (Appendix D) provides the correlation matrix. Table X provides a summary of the results of the all-grades regression. Table XI summarizes the F-ratios and R^2 change values for the nine sets of predictor variables employed. It should again be noted that these sets are not identical to those employed in the posttest regressions. Specifically, set four, Classroom Characteristic, includes Class Absence Rate, a variable not used in the posttest regression. In addition, the reading interest variable, which was a predictor in posttest analyses, is, of course, not now the criterion variable. Finally, Average Amount of Reading was dropped from the present analysis, since it is highly correlated with reading interest and was rated by teachers at the same time they were rating reading interest, leading us to suspect a halo effect. This would make the amount of reading variable one of questionable utility as a predictor of reading interest.

The final multiple correlation for the all-grades regression was .53, accounting for approximately 28% of the variance. The correlation is statistically significant.

Table XI shows that six of the nine sets of variables contribute significantly to the final regression equation. The first significant contributor was the pretest reading performance score. Classes with higher

TABLE X

READING INTEREST REGRESSION SUMMARY: ALL GRADES

Variable	Multiple R	R Square	RSQ Change	Simple R	F-Ratios		B	Beta
Pretest Score	.13962	.01949	.01949	-.13962	11.79***	†δ	-.17107	-.04875
Grade	.14428	.02082	.00132	.07209	SNS		.12483	.49372
Test Level Count	.14436	.0284	.00002	.03869	SNS		-.00698	-.01582
Class Size	.16415	.02695	.00610	-.06405	3.69	†	-.01123	-..9808
Mean Class Absence	.23473	.05510	.02815	.17734	17.02***	†δ	.00187	.01618
% Male	.25000	.06250	.00740	.08506	4.47*	δ	.00256	.05369
Holding Rate	.25007	.06253	.00003	-.03228	.018	†δ	.00459	.08197
Teacher Turnover	.29371	.08626	.02373	.15750	14.35***	†	.35026	.11951
Teacher Absence	.30808	.09491	.00865	.08417	5.23*		.40881	.07228
Percent Free Lunch	.38572	.14878	.05387	.31114	32.57***		.16717	.12621
Need for Help	.46280	.21418	.06540	-.37384	39.55 **	†δ	-.22213	-.21908
% Receive Help	.46593	.21709	.00291	-.04725	1.76	†δ	.07315	.05188
Health Barriers	.47326	.22397	.00688	.02999	SNS		-.05623	-.09197
Reading Problems	.47326	.22397	.00000	.13380	SNS	δ	-.01395	-.04117
Behavior Barriers	.47568	.22627	.00230	.13933	SNS		.02666	.05756
Math Specialist	.49864	.24864	.00032	.02456	0.19		.11388	.01753
Team Teaching	.50865	.25872	.00054	-.14660	0.32	δ	-.01790	-.01359
Open Space	.50812	.25818	.00954	-.09657	5.77*		-.09580	-.09464
Variability of Material Level	.49832	.24832	.02205	.21316	13.33***		.08385	.12279

Note: SNS means that the individual F was not calculated because the F for the set was nonsignificant.

*p < .05
**p < .01
***p < .001
† Significant in 1975-76
δ Significant in 1976-77

TABLE XI

READING INTEREST SCORE REGRESSION - ALL GRADES: CONTRIBUTION OF SETS

Set No.	Set Name	Number of Variables in set	F	1975-1976	1976-1977	R^2 Change
1	Reading Pretest	1	11.79***	†	δ	.019
2	Grade	1	.80			.001
3	Number of Test Levels	1	.01			.00002
4	Classroom Characteristics	4	6.30***	†	δ	.042
5	Teacher Characteristics	2	9.79***	†		.032
6	Percent Receiving Free Lunch	1	32.57***	†		.054
7	Reading Behavior	2	20.65***	†	δ	.068
8	Learning Barriers	3	1.85		δ	.009
9	Classroom Behavior	5	6.17***		δ	.051

*p < .05
**p < .01
***p < .001

† significant in 1975-76 year
δ significant in 1976-77 year

mean reading pretest scores have higher interest in reading. This result is consistent with those for both previous years of testing. Neither of the next two sets in the hierarchical analysis, Grade or Number of Levels, was significant, a finding also consistent with the previous two years.

The next set of variables to reach significance was Classroom Characteristics. Within that set two variables, Class Absence Rate and Percentage of Male Students, were significant. Consistent with prior years, the Class Absence Rate variable showed that classes with high absence rates have low levels of reading interest. The sex composition variable showed that classes with a higher percentage of male students have lower levels of interest in reading. This result was consistent with data for the 1976-77 year, but not with 1975-76. The Class Holding Rate variable, which had been significant in two previous years, was not significant in the 1977-78 data.

The significance of the Teacher Characteristic set was reflective of individually significant effects of both variables in that set. Low teacher turnover rates and low teacher absence rates are both predictive of high reading interest. The variables in this set should, however, be interpreted with caution. This is because the Teacher Turnover Rate variable was significant only in the 1975-76 data, while Teacher Absence Rate was not significantly related to reading interest in either of the previous years.

The next two sets to reach significance were Percentage of Students Receiving Free Lunch and Reading Behavior. The single-variable Percent Free Lunch set indicated that those classes having a small percentage of students who received free lunch may tend to have higher reading interest. Since this was not significant in either of the previous two years, its significance here should be interpreted with caution until it can

be replicated in data for future years. The Reading Behavior set contained one individually significant variable, the amount of extra reading help needed by the class. Classes needing larger amounts of help had lower interest in reading.

The final significant set was Classroom Behavior. Of the five variables in this set, three were significant. The first was the standard deviation (SD) of test levels for the class. This is our measure of class diversity of reading levels. The result obtained here indicates that classes with less reading level variability have higher interest in reading. The variable has not been significant in the previous two years and needs to be studied further in data for future years. The second significant variable was that indicating whether a given class was taught in an open-space or self-contained classroom. Results indicate that students taught in self-contained classrooms have higher interest in reading. Again, caution is needed until this variable can be replicated in future years.

b. Discussion and Comparison. The six sets of variables significant in the 1977-78 data have all been significant in one or both of the previous years, indicating a substantial degree of consistency across years at this level of analysis. Cutting across these six sets, a total of ten individual variables were significant in the all-grades analysis for 1977-78. The current analyses confirmed, first, our earlier finding that classes with higher average pretest scores have higher interest in reading. Here, the reading pretest score is not simply a control variable. Rather, its significance is reflective of a possible important underlying causal relationship to reading interest. Specifically, it may well be that

students who have relatively high levels of reading skills at the beginning of the year maintain, as a result, higher levels of interest in reading. Alternatively, a student who has high interest in reading may, as a result, read more and thereby improve his or her reading skills, resulting in a high pretest performance score. A further possibility is that interest and performance build cyclically on each other. The student who is interested in reading reads more, becomes more accomplished, and maintains the interest in reading, which leads to further performance increments. In any case, it would seem important to do whatever is possible to increase the interest the students have in reading and to apprise teachers of the importance of the interest variable as it relates to reading performance.

The classroom characteristics significant in the present study were Class Absence Rate and Percentage of Male Students. The first of these was significant in both 1975-76 and 1976-77 data. The fact that lower absence rates are predictive of higher reading interest is not surprising. The most likely interpretation would be that a student who is interested in reading is more likely to attend on a regular basis, although an alternative possibility is that a regular attendance actually increases reading interest. In either case, the predictive reliability of the Class Absence variable in the reading interest analysis indicates one more reason why class absences should be reduced as much as possible. The sex distribution variable (Percent Male Membership) suggests that classes with lower percentages of males may tend to exhibit higher reading interest levels. This finding is consistent with the 1976-77 data and may reflect the sociocultural encouragement of females to be more interested in academic pursuits and males less interested. It should be noted that the Class Holding Rates variable in the Classroom Characteristic set, though

not significant in the present data, was significant in both 1975-76 and 1976-77. It is therefore a variable that is worthy of further study in future years.

The stability of the teacher in the classroom is reflected in the significant Teacher Characteristics set. Both low Teacher Turnover rate and Teacher Absence rate are predictive of higher interest in reading. Thus, the more consistently the regular teacher is present in the classroom, the higher are likely to be the levels of reading interest. Results for these two variables, combined with the Class Absence Rate variable above, may suggest that a stable, consistent classroom in which most pupils and the assigned teacher attend regularly throughout the year are classes in which greater reading interest develops or is maintained. As was noted above, however, the two teacher variables, particularly the Absence Rate, will need to be further studied in future years before final determinations can be made.

The Percent Free Lunch variable, our socioeconomic indicator, has not ordinarily been significant in the all-grades analyses. Here it was significant in the expected direction, those classes with a higher percentage of free lunch students and therefore presumably a lower socioeconomic level showing lower reading interests. Among the most consistently significant sets of variables has been that dealing with reading behavior. In previous years both the amount of need for extra help and the amount of extra help received have been significantly predictive of reading interest. In the 1977-78 data only the first of these variables is significant, indicating that classes having a high need for extra help have relatively low levels of reading interest. These results have some important practical implications. If students are grouped in such a way that one class has a high need of extra help and another a low need, it

may be well to consider the alternative of grouping classes in such a way that no one classroom contains a high proportion of students likely to need extra help in reading. While this would have the effect of lowering reading interest somewhat in the classes having little need for extra help, it would provide at least a modicum of interest across all classes at a given grade level.

The final significant set in the present analysis, Classroom Behavior, is one in which we must, for the moment, have less confidence than in the other sets. This is because the set was significant in only one of the two previous years and, in that analysis, only one of the variables found to be significant here was statistically reliable. The two variables found significant only in the 1977-78 data were the standard deviation of first reading material and level and the type of classroom (open-space vs. self-contained). There is no difficulty in understanding the meaning of the reading material variability factor. It simply suggests that classes with a greater diversity of reading material level tend to have lower reading interest. The classroom type variable shows that students in self-contained classrooms have higher interest levels than those in open-space classrooms. This provides perhaps one more indication that the open-space variable should be investigated further.

3. The Analysis of Primary Grades

The means and standard deviations for predictor and criterion variables are provided in Table I (Chapter IV).

a. Results. Tables XII and XIII provide the regression summary for variables and sets in the analysis of primary grade data, and Table B (Appendix E) provides the correlation matrix.

The final multiple correlation of .54 accounted for approximately 29% of the variance. Four sets of variables significantly contributed to the prediction of reading interest scores in the primary grades. They were Teacher Characteristics, Percent Recieving Free Lunch, Reading Behavior, and Classroom Behavior.

The Teacher Characteristics and Percent Free Lunch Sets showed results consistent with those for the all-grades analysis reports above. Both Teacher Turnover rate and Teacher Absence rate within the Teacher Characteristics set were significantly predictive of reading interest. Greater classroom stability, reflected in low teacher turnover rates and low teacher absence rates, were associated with high levels of reading interest. The Percent Free Lunch variable indicated that primary grade classes with low percentages of free lunch students, and hence higher socioeconomic level, had higher interest in reading.

The remaining two sets of variables also showed results consistent with those in the all grades analysis. Within the Reading Behavior set, the average amount of need for extra help was significant, indicating that classes with greater need for help have lower interest in reading. Within the final significant set, Classroom Behavior, the significant variable was Standard Deviation of Reading Material Level. As in the all-grade analysis, the variable indicates that classes with less variation in reading

TABLE XII

READING INTEREST REGRESSION SUMMARY: PRIMARY GRADES

Variable	Multiple R	R Square	RSQ Change	Simple R	F-Ratios		B	Beta
Pretest Score	.10342	.01070	.01070	-.10342	SNS	†δ	-.08537	-.02255
Grade	.11547	.13334	.00264	.00208	SNS		.11456	.21711
Test Level Count	.13321	.01775	.00441	-.06297	SNS	†	-.00081	-.00146
Class Size	.15248	.02325	.00550	-.08405	SNS	†	-.01605	-.13820
Mean Class Absence	.19511	.03807	.01482	.13581	SNS		-.00202	-.01720
% Male	.19582	.03834	.00027	-.00019	SNS	δ	-.00404	-.08441
Holding Rate	.19793	.03917	.00083	-.00882	SNS	δ	.00379	.06997
Teacher Turnover	.26329	.06932	.03015	.17557	9.54**		.42582	.15515
Teacher Absence	.29441	.08668	.01736	.11535	5.49*		.77029	.11407
Percent Free Lunch	.35673	.12726	.04058	.27076	12.83***		-.28254	-.27564
Need for Help	.46401	.21531	.08805	-.40004	27.85***	†δ	-.28254	-.27564
% Receive Help	.46940	.22034	.00503	-.09036	1.59	δ	.12962	.08832
Health Barriers	.48421	.23446	.01412	-.02263	SNS		-.09026	-.15341
Reading Problems	.48810	.23824	.00378	.17442	SNS		.01525	.04627
Math Specialist	.51711	.26740	.00000	.00000	0.00	δ	.00000	.00000
Team Teaching	.52149	.27195	.00000	-.07089	0.00	δ	.02407	.01665
Open Space	.52149	.27195	.00455	-.03721	1.44		-.08911	-.08701
Variability of Material Level	.51711	.26740	.02855	.13784	9.03**		.12274	.17069

Note: SNS means that the individual F was not calculated because the F for the set was nonsignificant.

*p < .05
**p < .01
***p < .001
† Significant in 1975-76
δ Significant in 1976-77

TABLE XIII

READING INTEREST SCORE REGRESSION - PRIMARY GRADES: CONTRIBUTION OF SETS

Set No.	Set Name	Number of Variables in set	F	1975-1976	1976-1977	R^2 Change
1	Reading Pretest	1	3.38	†	δ	.011
2	Grade	1	.83			.003
3	Number of Test Levels	1	1.69	†		.004
4	Classroom Characteristics	4	1.69	†	δ	.021
5	Teacher Characteristics	2	7.51***			.048
6	Percent Receiving Free Lunch	1	12.83***			.041
7	Reading Behavior	2	14.72***	†	δ	.093
8	Learning Barriers	3	1.95			.019
9	Classroom Environment	5	3.54**			.056

*p < .05
**p < .01
***p < .001

† significant in 1975-76 year
δ significant in 1976-77 year

material level show higher interest in reading. However, the variable was not significant in either the 1975-76 or the 1976-77 data and should therefore be interpreted with caution.

b. Discussion and Comparison. Results of the regression analysis of reading interest in primary grades showed that four major variable sets were significant. These included Teacher Characteristics, Percent Receiving Free Lunch, Reading Behavior, and Classroom Behavior. Within the Teacher Characteristic set, results indicated that low teacher turnover rate and low teacher absence are associated with high reading interest. Classes with a lower percentage of students receiving free lunch and those with a lower need for extra help in reading were also likely to be higher in interest. The implications of all four of these important variables have been discussed in connection with the all-grades analysis above, and, since all results are directionally consistent with that earlier analysis, there is no need to repeat the discussion here.

The final significant set, Classroom Behavior, yielded reliable the Standard Deviation of Reading Material Level. Results from this variable was consistent with those seen in the all-grades analysis and do not require detailed repetition here. It should, however, be noted that the specific results seen here and not been seen in earlier years and hence will require replication before firm conclusions can be reached.

Just as notable as the similarities between the primary grades and all-grades analyses are the differences. In particular, it is notable that two of the sets significant in the all-grades analysis were not significant for the primary grades alone. The first of these was the reading pretest performance score which, in the all-grades analysis, is positively associated with reading interest. The second is the Classroom Characteristics set, which contains class absences and holding rates and sex distributions. Since the set was not significant, the individual variables were not separately tested for significance.

4. Analysis of Intermediate Grades

A separate regression analysis of data for the three intermediate grades paralleled that for the primary grades. The predictor variables were the same as in the two preceding analyses, and the regression was again of a hierarchical type.

a. Results. Table I (Chapter IV) contains the relevant means of standard deviations for the predictor and criterion variable. The correlation matrix appears in Table C (Appendix F). Table XIV and XV report, respectively, the regression summary for individual variables and the regression summary for sets.

The final multiple correlation of .58 accounted for approximately 33% of the variance.

It can be seen in Table XV that six of the nine sets of variables significantly predicted the reading interest criterion score in the intermediate grades analysis. The six significant sets were Pretest Reading Performance Score, Grade, Classroom Characteristics, Percent Receiving Free Lunch, Reading Behavior and Classroom Environment.

TABLE XIV

READING INTEREST REGRESSION SUMMARY: INTERMEDIATE GRADES

Variable	Multiple R	R Square	RSQ Change	Simple R	F-Ratios		B	Beta
Pretest Score	.15575	.02426	.02426	-.15575	7.02*	†δ	-.09630	-.02772
Grade	.21282	.04529	.02103	.08368	6.08*		.10966	.21365
Test Level Count	.23567	.05554	.01025	.08681	SNS	†δ	-.01346	-.03133
Class Size	.24959	.06230	.00676	-.06190	1.96		-.00879	-.07650
Mean Class Absence	.30747	.09454	.03224	.22642	9.32**	†δ	.00576	.05078
% Male	.36093	.13027	.03573	.20346	10.38**	†	.00938	.19544
Holding Rate	.36299	.13176	.00149	-.08372	.43		.00613	.10248
Teacher Turnover	.38343	.14702	.01525	.13777	SNS	†	.33796	.10795
Teacher Absence	.38845	.15090	.00388	.04799	SNS		.44149	.05965
% Free Lunch	.47242	.22318	.07229	.36217	20.91***	†δ	.25751	.19860
Need for Help	.51539	.26563	.04244	-.37429	12.28***	†δ	-.16470	-.16104
% Receive Help	.51885	.26921	.00358	-.02021	1.035	†	.05724	.04160
Health Barriers	.51984	.27023	.00102	.10695	SNS		-.01177	-.01831
Reading Problems	.52169	.27216	.00193	.09930	SNS		-.04565	-.12903
Behavior Barriers	.52332	.27386	.00171	.13019	SNS		.01746	.03841
Math Specialist	.54360	.29550	.00361	.04269	1.04		.26222	.05951
Team Teaching	.56484	.31905	.00420	-.23185	1.21		-.07281	-.06044
Open Space	.56112	.31485	.01935	-.16616	5.60*		-.10767	-.10819
Variability of Material Level	.54027	.29189	.01802	.29329	5.21*		.03788	.05879

Note: SNS means that the individual F was not calculated because the F for the set was nonsignificant.
*p < .05
**p < .01
***p < .001
† Significant in 1975-76
δ Significant in 1976-77

TABLE XV

READING INTEREST SCORE REGRESSION - INTERMEDIATE GRADES: CONTRIBUTION OF SETS

Set No.	Set Name	Number of Variables in sets	F	1975-1976	1976 1977	R^2 Change
1	Reading Pretest	1	7.02**	+	&	.024
2	Grade	1	6.08*			.021
3	Number of Test Levels	1	2.97	+	&	.010
4	Classroom Characteristics	4	5.51***	+	&	.076
5	Teacher Characteristics	2	2.77	+		.019
6	Percent Receiving Free Lunch	1	20.91***	+	&	.072
7	Reading Behavior	2	4.44*	+	&	.046
8	Learning Barriers	3	.45			.605
9	Classroom Environment	5	3.41**			.059

*p < .05
**p < .01
***p < .001

\+ significant in 1975-76 year
& significant in 1976-77 year

Results for the specific significant variables within sets were largely in accord with the all-grades analysis of the present data. The Pretest Reading Performance Score was, again, positively related to reading interest. The Mean Class Absence variable, as in the all-grades analysis, showed a negative relationship to the reading interest, such that higher class absences are predictive of lower reading interest. The Sex Distribution variable also accords with that in the all-grades analysis, showing that classes with a higher percentage of male students have lower interest in reading. The socioeconomic indicator, Percentage of Students Receiving Free Lunch, was also significantly predictive of reading interest. Again, those classes having lower percentage of students receiving free lunch had higher reading interest scores. Within the Reading Behavior set, the average amount of need for extra help with reading was once again significant. The classes with a high need for extra help show low interest in reading. Finally, the significant Classroom Environment set contained two individually reliable variable. The Standard Deviation of Reading Material Level was a significant predictor. Classes with less variability of material levels had higher reading interest. The second significant variable in this set showed that classes taught in self contained classrooms had higher reading interest than those taught in open-space settings. A final variable that was significant in the intermediate grades analysis was grade level. Results here suggest that reading interest declines as grade level increases across these three grades. The grade variable was not significant in the all-grades analysis and has not been significant in the analysis for previous years. It should therefore be interpreted with considerable

caution until it has been replicated in data for future years.

b. Discussion. Relatively little comment is needed concerning those variables that were consistent with the all-grades analysis, since their implications have been noted above. However, there are a number of additional points that should be made. First, Reading Pretest Score, again significant in this analysis, is one of the most consistent predictors of reading interest that we have. It has been significant in all three years of analysis at the intermediate grade level and in the all-grades analysis and should be considered to be an important variable for further study.

Three other variables that require particular further attention are Class Absence Rate, Percent Receiving Free Lunch, and Amount of Need for Extra Help. The first of these has been significant at the intermediate grade level in all three test years. Its implications have been discussed under the all grades analysis above, but it is emphasized here that the relationship between class absence rate and reading interest should be further studied at the intermediate grade level in an effort to see just how these two variables do relate and what kinds of action might be taken to reduce absences and increase reading interest levels. Similarly, the socioeconomic variable, Percent Receiving Free Lunch, has been significant in all three years at the intermediate grade level. While it may be that there is little appropriate action that could be taken with regard to this variable, it is important to be aware of its influence on reading interest. One possibility would be that teachers having high percentages of low socioeconomic children would need to make even greater efforts to instill reading interest in their classes. An alternative possibility is that different kinds of materials or different approaches to reading are necessary in classes with a predominance of low socioeconomic students, a point that might be explored with a reading specialist. The third

variable, Need for Extra Help, is totally consistent in predicting reading interest at both primary and intermediate grade levels and across all years of testing. The relationship of this variable to reading interest is very likely based in part on the fact that both variables derive from teacher ratings of individual students. Apparently, students who need considerable extra amounts of extra help are rated by their teachers as having lower interest in reading. While this may, in fact, be true, it may nevertheless represent a situation that can be dealt with. One implication of the relationship would be that a given student would be likely to need less help with reading if he or she became interested in the subject. Moreover, since the relationship is also significant at the primary grade level, the earlier this interest can be instilled, the better it would be for the student. Possibly this relationship, taken together with other data in the present study, suggests the need to consider placing greater early emphasis on increasing the child's interest in reading, as opposed to concentrating entirely on the development of technical reading skills.

VI. EVALUATIONS OF READING PROGRAMS

A. Methods

In this chapter evaluations of reading programs (as defined in Chapter IV D) are reported. The methods used were analysis of variance (ANOVA) and analysis of covariance (ANCOVA). Based on a set of selection criteria, types of programs were compared using the Posttest % Objectives Mastered on the Prescriptive Reading Test (PRT) as the dependent variable. These analyses, unlike the regression analyses discussed in the preceding chapter (Chapter V) used the student as the unit of analysis. That is, data for all students having the same type of reading program as defined in Chapter IV D) were grouped together regardless of whether or not students came from the same classes or even the same schools. For both sets of analyses, students were drawn from all classes at a given grade level, with the lower and upper dividing points being .5 or more and .4 or less to include multi-grade classes. That is, for example, all students in classes whose computed unweighted average class level was between 3.5 and 4.4 were classified as 4th grade students, regardless of their true grade level.

Two different analyses were done for each grade level. The first was a one-way analysis of variance (ANOVA) and covariance (ANCOVA) in which the independent variable was reading programs. The second analysis was a two-way analysis of covariance (ANCOVA) in which the independent variables were reading programs and whether or not the student received extra help in reading. The dependent variable in both sets of analyses was the percentage of posttest reading objectives mastered. The one-way analysis was done with the BMDP1V ANCOVA computer program; the second used the BMDP2V ANCOVA program.

For both sets of ANCOA (one-way, two-way), two covariates were used: pretest percentage of reading objectives mastered, and the coded reading level of the first reading materials (i.e., of the reading material used most frequently by the student). For both sets of analyses at a given grade level, the same reading programs and samples of students were used.

The same analyses were also done using a logit transformation of the two test variables. If the pretest or posttest percentage is converted to a proportion, p, the logit is: log (p + 1/6N) - log (1 - p + 1/6N) (Mosteller & Tukey, 1977). The N in the expression is the total number of objectives of the test. The descriptive data reported in the tables to follow are the percentage data. The significance tests reported are for the logit transformation analyses. The analyses of transformed pre and post measures gave results that were largely, but not entirely, similar to results for the untransformed variables. The logit transformation has the effect of spreading out the percentage scale increasingly the further the departure in either direction from 50% (or 0 logit). The transformation is folded, and variance is independent of position in the range.

The criteria for selecting types of reading programs to be compared for both sets of analyses were that the program had 50 or more students, and that losses (from missing variate or covariate data) left sample sizes of 30 or more. Programs (more correctly, types of programs) meeting those criteria at each grade level are identified by an asterisk in Table III (Chapter IV).

In the one-way analyses of covariance, if a test of the equality of regression slopes of the groups indicated heterogeneity (at the .05 level of significance), then tests were made of Neymann-Johnson regions

of significant difference between each pair of programs. Those tests were for posttest regressed on pretest. If slopes were homogeneous, but the test of the equality of adjusted group means was not significant at the 5% level of chance, no other statistical tests were interpreted. In the two-way ANCOVA, the test for interaction is the test of homogeneity of regression slopes. If the interaction was significant, no other tests were interpreted. Comparisons of pairs of programs used .05/n (n = number of pairs compared) as the criterion of significance of a difference between pairs.

The one-way set of analyses of covariance was made to test the null hypothesis that there was no difference in the effect of different reading programs on achievement. The two-way set was made to test the null hypotheses that there were no differences in effects of different reading programs and in receiving extra help in reading. There are assumptions underlying the use of analysis of covariance with the data reported here. One is that the effects of treatments are the same regardless of the starting levels (pretest of different groups). Starting levels of all groups were generally not equal. There are other assumptions involved (conceptual and technical) in the analyses. We thus believe that results should be viewed with some caution. It will be found, however, that analyses of data from the prior years ('75-'76 and '76-'77) will help to establish what conclusions can be drawn with greater and lesser confidence. That is because the main protective strategy employed was replication. If a similar pattern could not be replicated from one year to the next, conclusions were not drawn. Replication with non-random samples does not, of course, protect against all alternative hypotheses. Some of these are discussed as appropriate in the text to follow.

The reader is reminded that different students in a particular type of reading program (as defined) at a given grade level may have taken different levels of the PRT. Using percentages of objectives mastered as the measure, rather than the raw number of objectives, was the means employed to try to equalize scales for different test levels as much as possible short of conventional norming.

B. Comparison of the First Grade Reading Programs

Six types of reading programs met inclusion criteria:

- Basal-Linguistics/Phonics materials
- Basal-Programmed/Structured materials
- Basal-Language Experience materials
- Basal-Supplementary materials
- Basal-Special materials
- Linguistics/Phonics

The one-way analysis of variance test of unadjusted group means was significant ($F_{5,863} = 14.578$, $p = .0000$); the one-way analysis of covariance test of the equality of adjusted group means was also significant ($F_{5,861} = 4.653$, $p = .0003$). Regression slopes were homogeneous ($F_{10,851} = 1.331$, $p = .2088$). Table XVI shows for each group the unadjusted means and standard deviations, the pre-post difference of the unadjusted means, and the adjusted means for each reading program, based on the analysis of untransformed variables. It also shows the regression coefficients and the rank order of the unadjusted and adjusted post-test means for each group relative to the rest of the groups (1 = highest, 6 = lowest), based on the analysis of logit transformations.

Table XVII gives similar data for the two-way ANCOVA of reading programs and receiving or not receiving extra help in reading. In this analysis

TABLE XVI

STATISTICAL DATA FOR EACH READING PROGRAM COMPARED, GRADE 1

Program	Variable	Students	Mean	SD	Adjusted Mean	Regression Slope+	Rank of:+ Unadj. M.	Rank of:+ Adj.
Basal-Linguistic/Phonics		456						
	Pretest % Obj's Mastered		59.2	18.0		.68		
	Posttest %		85.6	14.4	84.3		2	2
	Reading Level*		3.6	1.4		.18		
	Pre-Post Diff:		26.4					
Basal-Language Experience		128						
	Pretest %		49.5	19.1		.27		
	Posttest %		78.8	15.4	83.1		6	5
	Reading Level		2.6	0.9		.24		
	Pre-Post Diff.		29.3					
Basal-Programmed/Structured		80						
	Pretest %		65.2	21.4		.46		
	Posttest %		85.4	13.5	83.7		3	3.5
	Reading Level		3.0	1.3		.37		
	Pre-Post Diff:		20.2					
Basal-Supplementary		105						
	Pretest %		52.0	17.1		.56		
	Posttest %		77.9	14.5	81.7		5	3.5
	Reading Level		2.5	1.4		.27		
Basal-Special		36						
	Pretest %		63.1	12.8		.42		
	Posttest %		82.9	10.8	81.6		4	6
	Reading Level		3.1	1.1		.36		
	Pre-Post Diff:		19.8					
Linguistic/Phonics		64						
	Pretest %		63.9	17.4		.65		
	Posttest %		91.1	10.1	87.7		1	1
	Reading Level		3.8	1.2		.16		
	Pre-Post Diff:		27.2					

*Reading Level is a coded variable, not the actual level.

+These numbers are based on logit transformation analysis; the rest of the data in the table are from the untransformed analysis.

TABLE XVII

COMPARISON OF READING PROGRAMS BY RECEIVING EXTRA HELP, GRADE 1*

Program	Extra Help	Number of Students	Pretest %	SD	Posttest %(Unadj)	SD	Posttest % (Adj)	Pre-Post Diff(Unadj)	Reading Level	SD	Slopes Pre	Slopes RL
Linguistic/Phonic	Yes	8	67.3	11.2	91.7	6.1	87.0	24.4	4.0	0.0	.33	
	No	56	68.4	18.2	91.0	10.6	88.0	27.6	3.8	1.3		2.17
Basal-Linguistic/Phonics	Yes	143	67.2	16.2	90.3	9.7	84.7	23.1	4.5	1.4		
	No	313	55.5	17.6	83.4	15.7	84.3	27.9	3.3	1.2		
Basal-Programmed	Yes	18	58.7	15.1	90.7	6.8	90.9	32.0	3.1	1.6		
	No	62	67.1	22.6	83.9	14.6	81.5	16.7	2.3	1.2		
Basal-Language Experience	Yes	33	53.4	21.5	81.8	17.4	84.7	28.4	2.8	0.8		
	No	95	48.5	18.2	77.7	14.6	82.5	29.2	2.5	.9		
Basal-Supplementary	Yes	44	48.3	16.5	79.9	13.2	83.7	31.6	3.8	1.6		
	No	61	54.7	17.2	76.5	15.3	81.1	21.8	2.1	1.1		
Basal-Special	Yes	12	67.9	9.6	88.1	8.7	83.2	20.2	4.0	0.0		
	No	24	60.7	13.8	80.4	10.9	80.7	19.7	2.7	1.2		
Average		869 (Total)	57.9		83.9		84.0	25.9	3.7			

Note: Reading Level (RL) is a coded variable, not the actual reading level recorded by the teachers.
*Includes students in multi-level classes whose grade level averages 1.4 or less.

there was a significant effect of reading programs ($F_{5,855} = 2.246$, $p = .048$), but not of receiving extra help ($F_{1,855} = 3.114$, $p = .078$). There was not a significant interaction ($F_{5,855} = 1.217$, $p = .299$).

Table XVIII shows the number of classes and number of students per class per program going into these analyses. The two analyses suggest that the type of reading program used at the first grade level did seem to have different effects in general. The fact that regression slopes were equal in the one-way ANCOVA suggests that relationships for that analysis were directly interpretable.

In the main comparison of reading programs (the one-way ANCOVA) at the first grade level, the Linguistic/Phonic only type of program was significantly more effective on the average than the other types, except the Basal-Linguistic/Phonics type, for the students in the sample analyzed. The other types of programs were not significantly different from each other in effectiveness, on the average with students throughout the pretest range.

The two-way analysis suggests that receiving extra help in reading did not seem to make a statistically significant difference in reading achievement. Thus, we can conclude only that this variable did not seem to have a constant incremental effect on the average. We can see in Table XVII that in most reading program groups, but not all, the students who were receiving extra help in reading were those with the lower average pretest scores.

It should be noted in Table XVI that the regression coefficients and rank orders come from the analysis of covariance using percentage objectives mastered pre and post transformed to logits. The other data in the Table (as well as all data in Table XVII) are from the analysis using

TABLE XVIII

COMPOSITION OF SAMPLES FOR ANALYSES OF READING PROGRAMS, GRADE I

Program	Number of Classes	Number of Students	Average Number of Students Per Class	Minimum Number of Students Per Class	Maximum Number of Students Per Class
Basal-Linguistics/Phonics	34	456	13.41	1	22
Basal-Programmed/Structured	7	80	11.43	1	23
Basal-Language Experience	12	128	10.67	1	21
Basal-Supplementary	9	105	11.67	1	22
Basal-Special	2	36	18.00	1	18
Linguistic/Phonic Only	4	64	16.00	1	22
Total Number of Different Classes, Students	62	869			

untransformed measures. The two analyses gave generally similar results. Thus, to avoid having to go through the process of reconverting the logits, we simply took percentages from the analyses using untransformed measures. The rank orders vary slightly. Therefore, there are times in which rank orders will be slightly different from the ranks that would be obtained from the percentages in the tables. All F-ratios are for the logit analyses. They are generally similar to those obtained from the analysis of untransformed data. One consequence of using the transformation is that it eliminated one interaction in the one-way ANCOVAS. Out of a total of 30 pairs of tests (all analyses), only 8 led to reversals of significance at the .05 level in comparing logit and percentage analyses of the same data.

Four of the same types of programs were compared both in 1976-77 and 1977-78 :

- Basal-Linguistic/Phonics
- Basal-Programmed/Structured
- Basal-Language Experience
- Linguistic/Phonics only

Performance relationships among these types of programs have been inconsistant, for the most part, over the two year period. The evidence over two years indicates that a Basal-Linguistic/Phonics approach produces results comparable to a Basal-Programmed/Structured approach. Otherwise, relationships have varied. For example, in 1976-77, Linguistic/Phonics only was the next to lowest ranking of the programs compared (based on adjusted posttest means); in 1977-78 it was the highest ranking. Thus, its performance relationships to other types of programs was reversed.

One hypothesis that might account for the marked change in the relative standing of the Linguistic/Phonics only type of reading program is that it is the result of sampling and selection bias. The sample sizes for the Linguistic/Phonics only type of program both years were small (26 in '76-'77; 64 in '77-'78). In '76-'77, the Linguistic/Phonics group had the lowest average pretest score of any of the groups compared. That was not the case in '77-'78. The pretest average of the sample was above the average for Grade 1. In '76-'77 the students making up the sample came from 10 different classes; there was a average of 2.6 students per class. In contrast, the '77-'78 sample came from 4 different classes for a average of 16 students per class (see Table XVIII). In effect it appears that the students in the '76-'77 sample were a few members of a number of classes for whom the teachers felt a Linguistic/Phonics program was appropriate; it was not the prevailing program in those classes. For the '77-'78 sample the opposite was the case; it was the prevailing program in most of the few classes involved. In that sense, the students in the '77-'78 sample was not so much a somewhat low pretest group scattered among a larger number of classes. They tended to be somewhat above average students and they were the products of a few teachers who predominantly concentrated on a Linguistic/Phonics approach with their classes. For that type of situation, the program was particularly effective. The specific teachers contributing to the 1977-78 sample were not the same ones contributing to the 1976-77 sample.

The evidence about the Basal-Linguistic/Phonic approach, compared to two of the other three approaches, has been mixed over the two year period. It was significantly more effective than a Basal-Language Experience approach in '77-'78; the opposite was the case in '76-'77. In this case the evidence supporting a possible selection bias hypothesis is less compelling

as far as sample distributions among classes is concerned. Both samples came from relatively large numbers of different classes, with at least moderately high average numbers of students per class. However, in '76-'77 the Basal-Language Experience sample had the highest average pretest scores of the groups compared; in '77-'78 it had the lowest, about 8% below the average score for Grade 1. Under those conditions, the Basal-Language Experience approach showed the largest average gain of any of the groups, but it was not sufficient to overtake the adjusted posttest mean of the Basal-Linguistic/Phonics group (see Table XVI). Indeed, the Basal-Language Experience type of program has produced the largest absolute gains of any of the types of programs compared in '76-'77 and '77-'78. In that respect, it has compared well with the other types of programs.

Four of the 12 different teachers contributing students to the Basal-Language Experience sample in 1977-78 were the same teachers (4 out of 13) contributing to it in 1976-77. The students from those 4 teachers made up 34% of the Language Experience sample in 1977-78, and 41% of it in 1976-77. The nature of the influence of that group of teachers on the performance characteristics of the Basal-Language Experience type of program each year is impossible to assess without more detailed analyses. Such relatively large proportions of students in the sample each year coming from the same teachers suggest that it would be appropriate to exercise caution in considering the representativeness of the samples with respect to different teachers.

A similar caution is applicable to the Basal-Programmed/Structured type of program. Thirty-eight percent of the sample of students in 1977-78, and 22% in 1976-77, had the same 2 teachers. They constituted 29% and 33% of the number of teachers contributing to the '77-'78 and '76-'77 samples, respectively.

The foregoing analyses concern two different issues that will recur in discussions of findings for grades 2-6. One concerns bias in samples by the characteristics of the students involved. The other concerns the representativeness of samples with respect to distributions of teachers employing different types of reading programs. Samples of students representing different types of reading programs are not the results of previous random assignments to treatments. They are the results of a number of prior selection decisions made by teachers, not by the Evaluation System. The intent of the Evaluation System is to make statements about the relative effectiveness of different approaches to teaching reading. Since groups representing different approaches are not formed by a true experimental design, with random assignment of students and teachers, there are threats both to the internal and external validity of results. The design used by the Evaluation System can not lead to conclusions about causes of the performance of students on the PRT. It can lead to statements about effectiveness <u>associated</u> with different types of programs, but not to statements about effectiveness <u>caused</u> by different approaches.

Clearly it would be desirable to know the types of students to which statements of associated effectiveness apply. Furthermore, it would be desirable to know to what population of teachers results pertain. If there is some reason to suspect that the students comprising a group classified as having had a particular type of reading program are somehow special, compared to students in general at a given grade level, one would want to be careful about generalizing conclusions. Similarly, if the group representing a type of approach to teaching reading consists of a significant proportion of students from one or two teachers, one would again want to exercise caution in interpreting the results and drawing

conclusions from them. The assumption otherwise is that if a group is made up of students from a number of teachers, its performance is less likely to be the result of the influence of a particular style, methodology or personality of particular teachers. There is thus more assurance that performance is associated with the type of program.

The analyses and results presented here do not lead to simple quantitative decision rules about threats to internal and external validity. We will, throughout, make a subjective assessment of potential threats by discussing data such as those just considered.

Analyses of the effect of receiving extra help in reading (the two-way ANCOVA) over a two year period have led to differing conclusions. In 1976-77, there was a significant interaction between receiving extra help and type of program. The conclusion was that the effect of receiving extra help depended on the type of reading program. That was not the case for 1977-78. The interaction was not significant, nor was the main effect.

C. Comparisons of Reading Programs at the Second Grade Level

Seven types of programs were selected for comparison at the 2nd grade level:

- Basal-Linguistic Phonics
- Basal-Programmed/Structured
- Basal-Language Experience
- Basal-Individualized Reading
- Basal-Supplementary
- Linguistic/Phonics-Basal
- Linguistic/Phonics

Table XIX gives the descriptive statistics for these groups. The one-way ANOVA test of equality of group means was significant ($F_{6,853}$ =

TABLE XIX

STATISTICAL DATA FOR EACH READING PROGRAM COMPARED, GRADE 2

Program	Variable	Number of Students	Unadjusted Mean	Unadjusted SD	Adjusted Mean	Regression Slope+	Rank of:+ Unadj. M.	Rank of:+ Adj. M.
Basal-Linguistic/Phonics		409						
	Pretest %		56.4	18.6		.69		
	Posttest %		79.8	14.0	82.4		3	3
	Reading Level		5.6	1.8		.14		
	Pre-Post Diff:		27.5					
Basal-Programmed/Structured		43						
	Pretest %		55.4	21.5	80.3	.39		
	Posttest %		80.3	12.5			4	4
	Reading Level		5.1	1.9		.28		
	Pre-Post Diff:		24.9					
Basal-Language Experience		80						
	Pretest %		57.8	18.8		.55		
	Posttest %		79.5	18.8	78.4		5	6
	Reading Level		5.1	2.03		.12		
	Pre-Post Diff:		21.6					
Basal-Individualized Reading		79						
	Pretest %		54.2	20.4		.55		
	Posttest %		77.3	19.3	76.1		6	7
	Reading Level		5.9	1.8		.27		
	Pre-Post Diff:		23.1					
Basal-Supplementary		124						
	Pretest %		45.8	18.2		.75		
	Posttest %		73.0	18.8	76.4		7	5
	Reading Level		5.3	1.8		.22		
	Pre-Post Diff:		27.2					

TABLE XIX CONTINUED

STATISTICAL DATA FOR EACH READING PROGRAM COMPARED, GRADE 2

Program	Variable	Number of Students	Unadjusted Mean	Unadjusted SD	Adjusted Mean	Regression Slope*	Rank of:+ Unadj. M.	Rank of:+ Adj. M.
Linguistics/Phonics		95						
	Pretest %		50.0	16.8		.87		
	Posttest %		84.0	15.6	87.2		2	2
	Reading Level		4.5	1.9		.20		
	Pre-Post Diff:		34.0					
Linguistics/Phonics-Basal		35						
	Pretest %		57.2	15.5		.88		
	Posttest %		88.5	12.6	90.9		1	1
	Reading Level		3.5	1.6		-.23		
	Pre-Post Diff:		31.3					

*Reading Level is a coded variable, not the actual level.
+These numbers are from the logit transformation analysis; the rest of the data in the table are from the untransformed analysis.

-99-

11.454, p = .0003). The ANCOVA test of adjusted posttest means was significant ($F_{6,856}$ = 17.381, p = .000). The regression slopes of the groups were not homogenous ($F_{12,844}$ = 2.120, p = .014).

Table XX shows descriptive data for the two-way analysis of programs. In the two-way ANCOVA of whether or not the students were receiving extra help in reading, there was a significant effect of reading programs ($F_{6,849}$ = 18.888, p = .000), and of extra help ($F_{1,849}$ = 6.453, p = .011). The interaction, however, was also significant ($F_{6,849}$ = 5.908, p = .000). Table XXI gives sample composition information.

The analyses suggest that there were differences in the effectiveness of different types of programs, but that the relationships were complex. That is, different programs seemed to be affecting similar types of students (defined by pretest scores) differently. Figure 1 shows the regression lines of each program for posttest regressed on pretest. The axes are for the logit scale. Figure 1 also shows the regions of significant differences between slopes for designated pairs of types of reading programs. Those regions are indicated by bars, with the programs compared designated by numbers. Figure 1 shows graphically what is meant by a complex relationship of differences among programs, or by an interaction.

Using the logit transformed data, Neymann-Johnson regions of significance (at the familywise .05 level) between regression slopes were computed for each pair of type of program. Figure 1 shows clearly that several different types of programs produced significant interactions with the other types of programs insofar as regression of posttest on pretest is concerned (analyses of regressions of posttest on pretest and reading level of most used material were made, but will not be discussed in this report). Since it will be found that similar analyses, figures, and interpretations will be given for Grades 3 and 4, it is worth examining Figure 1

TABLE XX

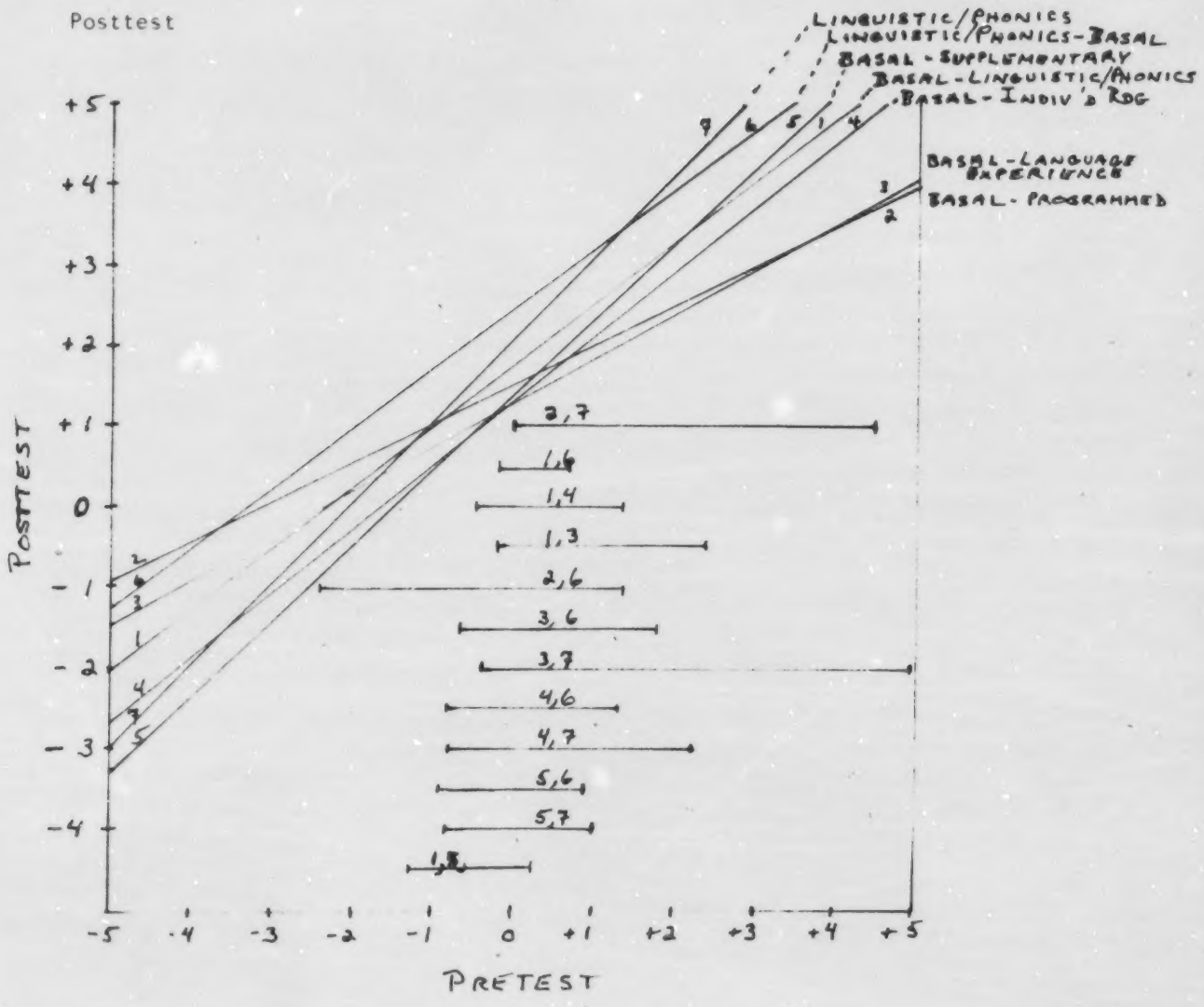
COMPARISON OF READING PROGRAMS BY RECEIVING EXTRA HELP, GRADE 2*

Program	Extra Help	Number of Students	Pretest %	SD	Posttest %(Unadj)	SD	Posttest % (Adj)	Pre-Post Diff(Unadj)	Reading Level	SD	Pre
Basal-Linguistic/Phonics	Yes	172	55.4	18.7	85.6	12.2	83.3	30.2	6.2	1.6	.40
	No	237	57.1	18.5	82.7	15.1	81.6	25.6	5.2	1.9	
Basal-Programmed/Structured	Yes	12	68.5	16.3	85.7	17.8	78.7	17.2	5.9	1.8	
	No	31	50.4	21.3	78.2	11.9	80.9	27.8	4.7	1.8	
Basal Language Experience	Yes	45	54.8	18.8	79.2	12.8	77.9	24.4	5.8	1.6	
	No	35	61.9	18.4	79.8	12.2	79.1	17.9	4.1	2.1	
Basal/Individualized Reading	Yes	54	54.4	22.4	78.6	18.5	76.6	24.2	6.3	1.8	
	No	25	53.6	15.5	74.6	20.9	75.0	21.0	5.2	1.5	
Basal-Supplementary	No	43	52.9	19.5	76.1	17.9	75.5	23.2	5.9	1.4	
	Yes	81	42.0	16.4	71.3	19.1	76.9	29.3	5.0	2.0	
Linguistic/Phonics	Yes	36	52.7	15.9	84.3	13.8	85.7	31.6	4.9	2.1	
	No	59	48.4	17.3	83.7	16.7	88.2	35.4	4.3	1.8	
Linguistic/Phonics-Basal	Yes	11	58.4	14.1	94.3	6.4	96.8	35.9	3.2	1.7	
	No	24	54.4	18.6	75.8	13.7	21.4	4.2	1.1		
Average		865 (Total)	54.1		81.3		81.0	27.2	5.3		

Note: Reading Level (RL) is a coded variable, not the actual reading level recorded by the teachers.
*Includes students in multi-level classes whose grade level averages 1.5 to 2.4.

Figure 1

Regions of Significant Difference Between Pairs of Reading Programs, Grade 2



Note:

Because of the large number of interactions, bars are used to depict regions of significant differences or of non-significance (NS) between the regression slopes of reading programs. Numbers designate the pairs of programs to which the bars pertain.

TABLE XXI

COMPOSITION OF SAMPLES FOR ANALYSES OF READING PROGRAMS, GRADE 2

Program	Number of Classes	Number of Students	Average Number of Students Per Class	Minimum Number of Students Per Class	Maximum Number of Students Per Class
Basal-Linguistics/Phonics	27	409	11.05	1	27
Basal-Programmed/Structured	3	43	14.33	8	18
Basal-Language Experience	10	80	8.00	1	16
Basal-Individualized Reading	12	79	6.58	1	22
Basal-Supplementary	13	124	9.54	1	18
Linguistics/Phonics	13	95	7.31	1	16
Linguistics/Phonics-Basal	4	35	8.75	1	16
Total Number of Different Classes, students	78	865			

in some detail before interpreting the results.

First, the scales for pretest and for posttest pertain to the logit transformation and cover only the approximate range of possible values for it. A logit value of 0 is equivalent to 50% objectives mastered (either pre or post). A logit of around +5 or -5 is equivalent to around 100% and 0% objectives mastered, respectively. We say "around" because the exact logit values for those percents depend on the test level, since the logit is based on the number of objectives and different test levels have different numbers of objectives. A logit of +1 or -1 is approximately equivalent to 75% or 25% respectively. A logit of +2 or -2 is approximately equivalent to 90% or 10% respectively, and a logit of +2.8 is roughly equal to about 94%. Thus, in Figure 1, the fact that the regression line for Linguistic/Phonics reaches a maximum positive value of about +2.8 on the pretest scale means that there were no students in the sample for that type of program who had a pretest score of above about 94% reading objectives mastered. In fact, the highest pretest score in that group was 85.7%. The same, of course, might have been true for the other types of programs, but the linear projections of the regression lines did not reach +5 on the posttest scale until further along the pretest line. The regression tests and regression lines are based on the assumption of linear regression. Slopes were plotted simply by fairing lines through three points whose values were determined from the regression equation for each program regardless of the actual distribution of pre and posttest scores.

For further discussion for the moment, let us refer to types of programs by the numbers shown in Figure 1 as a matter of convenience. Note that the bars under the group of regression lines in fact demarcate a region or zone along the pretest axis in which the regression lines for a

pair of programs differ significantly. The bars could also demarcate a region or zone within which the slopes of a pair of programs do not differ significantly, although this case does not occur in Figure 1. The brackets for the comparison of programs 2 and 7, for example, are of the first type. Thus, for programs 2 and 7, there is a significant difference between regression lines for pretest values of about 0 or better, out nearly to the end of the pretest range (+5.0). The two lines do not differ significantly over the rest of the range of the pretest scale (i.e., from less than 0 to -5). Actually, as we can see in Figure 1, and as we have stated before, the region of significance can only extend out to about +2.8 on the pretest range since that is where the program 7 line reaches the maximum value possible (+5.0) on the posttest scale. We also, of course, cannot tell from Figure 1 whether there were any students in the sample for program 2 who had any pretest scores of greater than +2.5. In fact there were not. The vast majority of students in all programs in fact fall between $\pm$2.5, however.

Note that for most of the pairs of programs in Figure 1, no bracket of either type is shown. If no bracket is shown for a pair of programs (e.g., 1, 2, or 2, 3, in Figure 1), it means that there was no significant difference between the slopes of those two programs over the whole possible range of the pretest variable as measured in logits. This is the case for most of the pairs of programs in Figure 1. As a general note, if the test of differences between means had been significant, but the test for homogeneity or equality of slopes not significant, the lines would appear separated but essentially parallel.

Before interpreting the results shown in Figure 1, let us examine one more important fact in it. For the convenience of the reader, the program numbers have been written on the appropriate regression lines at both ends

of the pretest (or X-) axis. Note that at the lower (left) end of the pretest scale, program 7 is next to the bottom, while at the upper (right) end it is at the top. The regression line of program 7 crosses the lines of the other programs, at different points on the pretest scale. When two regression lines cross within the range of the pretest and posttest variables, they are called disordinal. Thus, the regression slopes of program 7 and most of the other programs in Figure 1 are disordinal. Note, incidentally, that the lines for programs 1 and 4 are not disordinal; they do not cross within the range of the pre and posttest variables. The line for program 1 is higher (though not significantly so) than that for program 4 over the whole range.

The significant disordinal relationships in Figure 1 tell us something very important. They tell us that, on the average, students in program 7 who did well on the pretest did better than expected on posttest compared to similar students in some of the other types of programs. The numbered horizontal bars in the figure show the approximate regions of significant difference between program 7 and certain of the other types of programs. The regions of significance (or the bracketed regions of non-significance which will be seen in subsequent figures) also tell us that students in program 7 who scored below the middle of the pretest range on pretest on the average did no better or worse on posttest than similar students in the other programs. For such students it made no difference what type of reading program they were in; all seven types of programs were equal in effectiveness for the samples of students involved in the analysis. It is important in any case to note that when we say "better than expected" or "poorer than expected," we mean in relation to the performance of the specific samples of students in the particular type of

programs used in this analysis. We do not mean in relation to an external standard. The comparisons are of the programs and samples of students with each other. That is why final conclusions about the relative effectiveness of types of reading programs must be qualified.

Figure 1 indicates that for students who scored average or above on pretest, the Linguistic/Phonics only type of program was more effective on the average than Basal-Programmed/Structured, Basal-Language Experience, or a Basal-Individualized Reading type of program. There is indication of a significant difference in the pretest mid-range between Linguistic/Phonics only and Basal-Supplementary. Note, however, that the slopes of those two programs are not disordinal, and not widely separated over most of the range. The difference is probably significant in the mid-range because that is where the heaviest concentration of scores occurred. The result, in effect, is probably due to the distribution of scores along the pretest range.

Figure 1 also indicates that for students who scored average or below on pretest, the Linguistic/Phonics-Basal type of program was more effective, on the average, than a Basal-Linguistic/Phonics, Basal-Individualized Reading or Basal-Supplementary type of program. It appeared to be significantly more effective than a Basal-Programmed/Structured or Basal-Language Experience type only over the lower or middle pretest range, but that again is probably a result of sample distributions. The regression lines for those comparisons diverge from left to right. The non-significant regions would probably occur only at the extreme lower end of the pretest scale if there were sufficient sample sizes along the whole range (assuming regression to be linear).

Finally, Figure 1 indicates that a Basal-Linguistic/Phonics type of program was significantly more effective, on the average, with students

who scored average or above on pretest than a Basal-Language experience type. There is also indication that it was more effective than a Basal-Individualized Reading type with students who scored average or below on pretest.

Figure 1, in sum, makes graphically clear what is meant by calling relationships among programs complex. In general it appeared that the Basal-Programmed/Structured and Basal-Language Experience types of programs were less effective, on the average, than programs using a linguistic/phonics material and emphasis with average or above students.

How do these relationships compare to those found in 1976-77? The most obvious change comes from the finding in 1976-77 that there were no significant differences and interactions among the programs compared. However, there also have been some consistencies. Thus, over the two year period, there have been no significant differences in relative effectiveness among the following:

- Basal-Linguistic/Phonics and: Basal-Programmed/Structured, Basal-Supplementary, Linguistic/Phonics only, Linguistic/Phonics-Basal;
- Basal-Programmed/Structured and: Basal-Linguistic/Phonics, Basal-Language Experience, Basal-Individualized Reading, Basal-Supplementary;
- Basal-Language Experience and: Basal-Individualized Reading, Basal-Supplementary, Basal-Programmed/Structured;
- Basal-Individualized Reading and: Basal-Programmed/Structured, Basal-Language Experience, Basal-Supplementary, Linguistic/Phonics-Basal;

- Basal-Supplementary and: Basal-Linguistic/Phonics, Basal-Programmed/Structured, Basal-Language Experience, Basal-Individualized Reading;
- Linguistic/Phonics only and: Basal-Linguistic/Phonics, Linguistic/Phonics-Basal;
- Linguistic/Phonics-Basal and: Basal-Individualized Reading, Linguistic/Phonics only.

Over the past two years there have been some consistent patterns in relationships with respect to differences in performance effectiveness. We here apply the decision rule formulated in the report: Evaluation of Outcomes, 1976-77: An Evaluation System Report on Reading Programs and Reading Achievement, Part II - A Technical Report, page 105. That rule is that no conclusion be drawn, even tentatively, unless the same performance relationship of two or more programs can be shown over at least two years, and unless in at least one of those years a statistically significant difference can be shown.

By that rule, the following similar relationships have been found in the '76-'77 and '77-'78 analyses. There is evidence that a linguistic/Phonics-Basal approach tends to be more effective than a Basal-Linguistic/Phonics approach, a Basal-Programmed/Structured approach, a Basal-Language Experience approach and a Basal Supplementary approach. That evidence, however, must be tempered by the fact that in both years the sample sizes in the Linguistic/Phonics-Basal type of program were small (35 students) and the students came from a small number of different classes (see Table XXI). The results, therefore, could reflect the effectiveness of a few teachers using a type of reading program with which they were particularly comfortable and experienced. Fifty percent of the teachers

were the same both years, accounting for 66% of the sample in 1977-78 and 38% in 1976-77. Both years students in the group have scored above average in pretest.

Changes in the performance relationships among the type of programs compared over the past two years are described next. Such changes (e.g., evidence suggesting reversals of effectiveness relationships) leave as yet in doubt just how such types of programs are likely to perform in relation to each other.

- Basal-Linguistic/Phonics and: Basal-Language Experience, Basal-Individualized Reading;
- Linguistic/Phonics only and: Basal-Programmed/Structured, Basal-Language Experience, Basal-Individualized Reading, Basal-Supplementary.

The Linguistic/Phonics only group both years had a pretest average that was low among the groups concerned. Four teachers have contributed heavily to the sample size each year (44% in '77-'78, 73% in '76-'77. The Basal-Programmed/Structured group has also contained relatively large proportions of students. One teacher contributed 19% of the sample in '77-'78 and 33% in '76-'77. Thus there is reason to believe that the performance of these two types of programs may be particularly influenced by a few particular teachers.

It is unfortunate that there was an insufficient sample size to include the Basal only type of program in the '77-'78 comparisons.

Over the three year period (1975-78), four types of programs have been included in all three analyses: Basal-Linguistic/Phonics, Basal-Programmed/Structured, Basal-Supplementary, and Linguistic/Phonics only. Of the six pairs of comparisons over the three year period, only one has

shown a consistent performance relationship. The Basal-Linguistic/Phonics and Linguistic/Phonics only types of program have consistently been comparable in effectiveness.

From '76-'77 to '77-'78, the effect of receiving extra help in reading changed from no significant effect to indication of an effect that depended on the type of program involved. It can be seen in Table XX that, as has been the case in the past, the average pretest scores of groups of students receiving extra help in reading were lower than those not receiving it, depending on the type of program involved.

D. Comparisons of Reading Programs at the Third Grade Level

Seven types of program were compared:

- Basal
- Basal-Linguistic/Phonics
- Basal-Programmed/Structured
- Basal/Supplementary
- Basal-Language Experience
- Basal-Individualized Reading
- Linguistics-Phonics

Descriptive statistics for the programs are given in Table XXII. The differences between the unadjusted means, as well as between the adjusted means, were highly significant (respectively: $F_{56,929} = 8.756$, $p = .0007$; $F_{6,927} = 10.845$, $p = .000$). The regression slopes of the groups, however, were not homogeneous ($F_{12,915} = 3.359$, $p = .0000$). Thus, the analysis of covariance was not directly interpretable. An analysis of regions of significant difference between programs was needed and done.

TABLE XXII

STATISTICAL DATA FOR EACH READING PROGRAM COMPARED, GRADE 3

Program	Variable	Number of Students	Unadjusted Mean	Unadjusted SD	Adjusted Mean	Regression Slope+	Rank of:+ Unadj. M.	Rank of:+ Adj. M.
Basal		62						
	Pretest % Obj's Mastered		47.1	9.7		.13		
	Posttest %		86.8	14.1	84.4		1	2
	Reading Level*		7.5	1.5		.29		
	Pre-Post Diff:		39.7					
Basal-Linguistics/Phonics		385						
	Pretest %		43.5	16.7		.64		
	Posttest %		76.4	16.1	76.2		4	5
	Reading Level		7.2	1.9		.17		
	Pre-Post Diff:		32.9					
Basal-Programmed Structured		66						
	Pretest %		43.5	16.3		.79		
	Posttest %		73.1	17.0	73.3		6	7
	Reading Level		7.1	1.8		.19		
	Pre-Post Diff:		29.6					
Basal-Language Experience		33						
	Pretest %		39.5	16.6		.34		
	Posttest %		71.7	13.5	77.5		7	4
	Reading Level		5.7	2.8		.03		
	Pre-Post Diff:		32.2					
Basal-Individualized Reading		144						
	Pretest %		47.5	13.2		.26		
	Posttest %		82.2	11.7	79.8		2	3
	Reading Level		7.4	2.1		.25		
	Pre-Post Diff:		34.7					
Basal-Supplementary		201						
	Pretest %		1.4	13.6		.72		
	Posttest %		77.3	15.1	77.3		5	6
	Reading Level		7.5	2.1		.13		
	Pre-Post Diff:		35.9					
Linguistic/Phonic		45						
	Pretest		38.8	16.8		.61		
	Posttest %		77.5	22.1	85.2		3	1
	Reading Level*		5.1	2.1				
	Pre-Post Diff:		38.7			.37		

*Reading Level is a coded variable, not the actual level.
+These numbers are from the logit transformation analysis; the rest of the data in the table are from the untransformed analysis.

Table XXIII gives descriptive statistics for the two-way analysis of covariance for third grade level programs. In the two-way ANCOVA, there was a significant effect for reading programs ($F_{6,920} = 10.021$, $p = .000$), but there was not a significant effect for receiving extra help in reading ($F_{1,920} = .865$, $p = .353$). The interaction was significant ($F_{6,920} = 3.743$, $p = .001$). Table XXIV gives the sample compositions for Grade 3.

Figure 2 shows the regression lines of the types of programs compared. It also shows the regions of significant difference between pairs of programs (horizontal bars indicate regions and the numbers indicate the pair of programs to which the bar pertains).

Figure 2 depicts the results of the comparisons of programs for 1977-78. These may be summarized as follows:

· For students who scored below average on pretest, a Basal only type of program was associated with greater effectiveness than : a Basal-Linguistic/Phonics, a Basal-Language Experience, Basal-Programmed/Structured, and a Basal-Supplementary type; students in a Basal only type of program were comparable in performance with those in a Basal-Individualized Reading and a Linguistic/Phonics only type of program.

· For students who scored above average on pretest, a Linguistic/Phonics only type of program was associated with greater effectiveness than : a Basal-Linguistic/Phonics, Basal-Programmed/Structured, Basal-Language Experience, Basal-Individualized Reading, and a Basal-Supplementary type of program; the Linguistic/Phonics only type was comparable in effectiveness with a Basal only type with all students.

· The Basal-Individualized Reading type of program was more effective than a Basal-Linguistic/Phonics type and a Basal-Programmed/Structured

TABLE XXIII

COMPARISON OF READING PROGRAMS BY RECEIVING EXTRA HELP, GRADE 3*

Program	Extra Help	Number of Students	Pretest %	SD	Posttest %(Unadj)	SD	Posttest % (Adj)	Pre-Post Diff(Unadj)	Reading Level	SD	Slopes Pre	Slopes RL
Basal	Yes	37	47.0	13.4	85.6	8.0	84.5	38.6	7.8	0.8	.62	.17
	No	25	47.3	15.4	85.6	11.8	84.5	38.3	7.0	2.2		
Basal-Linguistic/Phonics	Yes	137	43.4	15.5	75.9	15.9	75.4	32.5	7.4	1.7		
	No	248	43.6	17.3	76.7	16.2	76.7	33.1	7.2	2.0		
Basal-Programmed Structured	Yes	18	44.5	8.8	84.4	9.8	79.3	39.9	9.0	1.0		
	No	48	43.0	18.4	68.9	17.3	71.1	23.9	6.4	1.5		
Basal-Language Experience	Yes	1	46.2	0.0	69.2	0.0	63.4	23.0	9.0	0.0		
	No	32	39.3	16.8	71.8	13.7	77.6	32.5	5.6	2.8		
Basal-Individualized Reading	Yes	75	47.2	12.4	85.4	9.4	80.9	38.2	8.3	1.7		
	No	9	47.8	14.1	78.6	13.0	78.7	30.8	6.4	2.0		
Basal-Supplementary	Yes	62	43.2	14.1	79.7	13.2	78.7	36.5	7.6	6.0		
	No	139	40.6	13.4	76.2	15.8	76.7	35.6	7.4	6.0		
Averages		936 (Total)	43.6		77.8		75.6	34.2	7.2			

Note: Reading Level (RL) is a coded variable, not the actual reading level recorded by the teachers.
*Includes students in multi-level classes whose grade level averages 2.5 to 3.4.

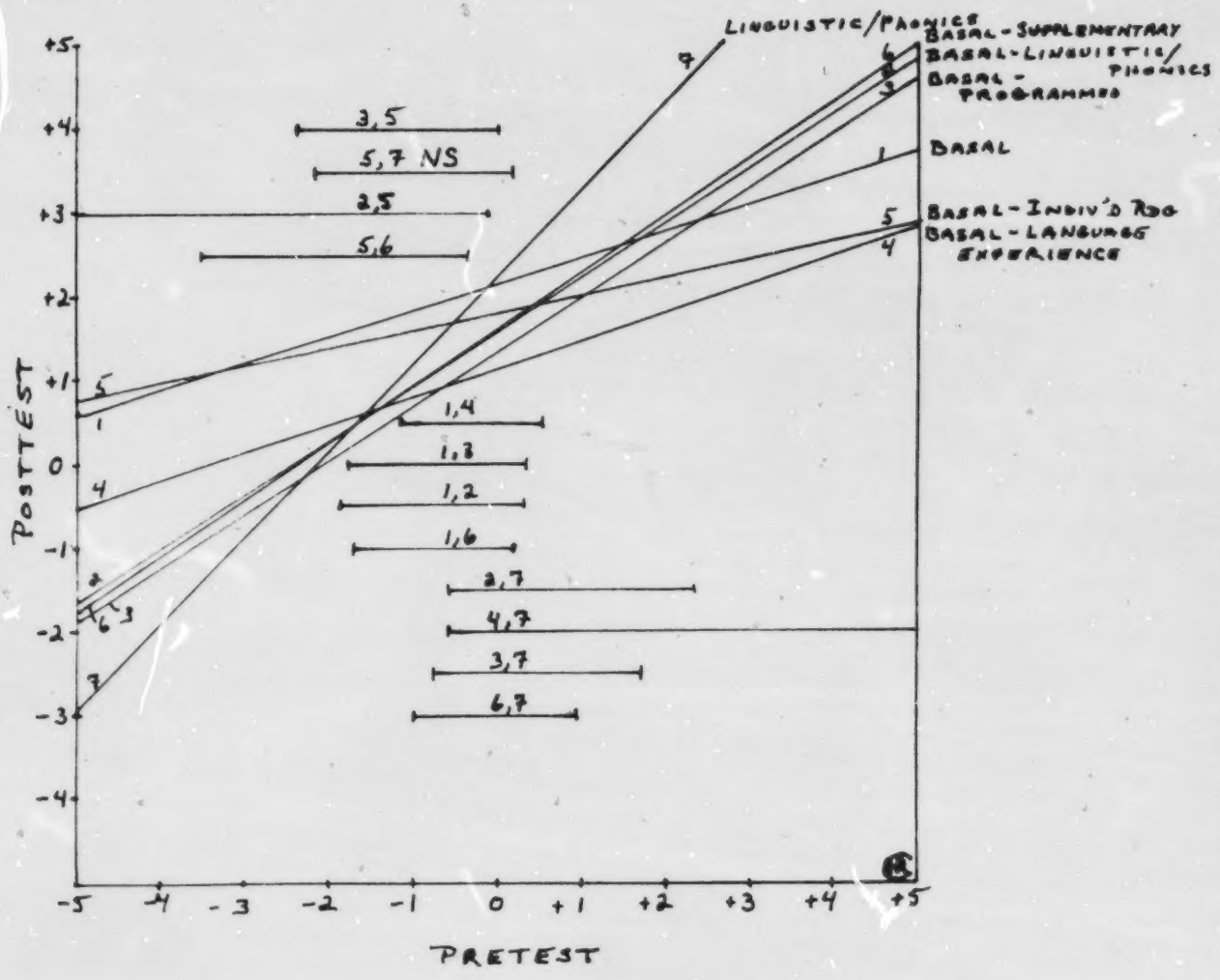
TABLE XXIV

COMPOSITIONS OF SAMPLES FOR ANALYSES OF READING PROGRAM, GRADE 3

Program	Number of Classes	Number of Students	Average Number of Students Per Class	Minimum Number of Students Per Class	Maximum Number of Students Per Class
Basal	8	62	7.75	3	15
Basal-Linguistics/Phonics	39	385	9.87	1	24
Basal-Programmed/Structured	7	66	9.43	3	15
Basal-Individualized Reading	18	144	8.00	1	20
Basal-Supplementary	17	201	11.82	2	21
Linguistics/Phonics	6	45	7.50	1	18
Total Number of Different Classes, Students	78	936			

Figure 2

Regions of Significant Difference Between Pairs of Reading Programs, Grade 3.



Note:

Because of the large number of interactions, bars are used to depict regions of significant differences or of non-significance (NS) between the regression slopes of reading programs. Numbers designate the pairs of programs to which the bars pertain.

type, with below average students.

· The Basal-Supplementary type was less effective than a Basal-Individualized Reading type of program and a Basal only type, with students scoring below average on pretest; the Basal-Supplementary type of program was comparable in effectiveness with a Basal-Linguistic/Phonic, and Basal-Language Experience, and a Basal-Programmed/Structured type.

· Except as noted above, the Basal-Language Experience type was comparable in effectiveness with: Basal only, Basal-Linguistic/Phonic, Basal-Programmed/Structured, Basal-Individualized Reading, and Basal Supplementary.

Compared to results for 1976-77, the 1977-78 results showed both consistencies and inconsistencies. These consistencies have occurred.

A Linguistic/Phonics only type of reading program has been associated with greater effectiveness than any of the other types* with students who score above average on pretest; there are indications (not statistically significant, however) that the Linguistic/Phonics only type of approach may be less effective than the other types compared with below average students.

A Basal-Individualized Reading type of program has appeared to be more effective than a Basal-Linguistic/Phonics type with students who score below average on pretest; it has been comparable in performance with a Basal only and a Basal-Language Experience type for all students.

*The Basal-Programmed/Structured type of program was not included in the 1976-77 evaluations. Thus, comparative statements about consistent or inconsistent results can not include that type. The types of programs compared both years were: Basal only, Linguistic/Phonics only, Basal-Programmed/Structured, Basal-Language Experience, Basal-Linguistic/Phonics, Basal-Individualized Reading, and Basal-Supplementary.

- A Basal-Language Experience type of program has been comparable in effectiveness to a Basal-Linguistic/Phonics, a Basal-Supplementary, a Basal-Individualized Reading, and a Basal-Programmed/Structured type.

- A Basal Supplementary type of program has been comparable in effectiveness to a Basal-Linguistic/Phonics and a Basal-Language Experience type of program.

The inconsistencies observed over a two year period have been in these comparisons:

- Basal only and : Basal-Linguistic/Phonics, Basal-Language Experience, Basal-Supplementary;
- Basal-Individualized Reading and Basal Supplementary

Performance relationships between these pairs of programs have not shown a consistent pattern over a two year period. Thus, it is not possible yet to draw conclusions about relative effectiveness.

Both in 1976-77 and in 1977-78, the Linguistic/Phonics only and the Basal-Language Experience groups were the samples with the two lowest pretest averages. They had the lowest average reading levels of the most used reading material, although that was not the case for the Linguistic/Phonics group in 1976-77. They both had sample sizes each year that were small compared with most of the rest of the groups. Thus, they came from relatively small numbers of different classes each year (see Table XXIV for 1977-78). The samples for the Basal-Language Experience did not, however, come from the same teachers both years. Therefore, if students in the Basal-Language Experience group were showing in part the effects of particular teachers, they were not the same 8 teachers each year. There is, as a consequence, some assurance about the stability and generalizability of the performance characteristics of that type of programs despite

It should be noted that the pretest variances of all groups were smaller in '77-'78 than in '76-'77. But there was also another marked change: pretest means were markedly lower (an average of 43.6% objectives mastered in the fall, compared to 62.8%). The pretest averages for Grades 1 and 2 were somewhat also lower (about 2% and 5% respectively) in '77-'78, but not nearly as low as those for Grade 3, and also Grades 4-6 (see, for example, Table II in Chapter IV). Posttest averages tended to be a little higher, but not dramatically (about 2%), than in '76-'77 for Grades 1, 2, 4, 5, and 6. The immediate consequence for Grades 3-6 was markedly higher gains in '77-'78.

The reasons for the drop in pretest scores, which start with Grade 3 are not known. It is possible, however, that whatever caused them could be a factor contributing to the apparent inconsistencies in performance relationships. It is possible, but not convincing. It would not account for inconsistencies found in Grades 1 and 2, which did not show the marked drop in pretest averages. Nor is it clear why it would affect some comparisons in Grade 3, but not the majority of them.

In neither 1976-77 and 1977-78 was there a significant effect of receiving extra help in reading, but there was a significant interaction of extra help and type of program. Indeed, that was also the case in 1975-76. In effect, in all three years the effectiveness of extra help appeared to depend on the type of reading program. That is a different situation from Grades 1 and 2 in which there was a significant interaction one year but not the other.

E. Comparisons of Reading Programs at the Fourth Grade Level

Five types of program met the selection criteria:

- Basal
- Basal-Linguistic/Phonics

- Basal-Individualized Reading
- Basal-Supplementary
- Linguistics-Phonics

Table XXV presents the descriptive statistics for each type of program. The test of differences of unadjusted means was significant, as was the test of adjusted means at the .05 level (respectively: $F_{4,775} = 4.386$, $p = .002$; $F_{4,773} = 2.091$, $p = .080$). The regression slopes were not homogenous ($F_{8,765} = 2.918$, $p = .003$).

Table XXVI gives descriptive statistics for the two-way analysis of covariance. In this analysis, the reading program effect was significant ($F_{4,768} = .699$, $p = .593$). The test of receiving extra help in reading was not significant $F_{1,768} = 1.700$, $p = .193$), nor was the interaction significant ($F_{4,768} = 1.967$, $p = .098$). Table XXVII gives the sample compositions of the different programs.

From the one-way ANCOVA it may be concluded that there appears to be a difference in effectiveness of the reading programs, but that, again, relationships are complex. Figure 2 shows the regression lines of the types of programs compared.

It may be seen in Figure 2 that two types of reading programs produced the significant interactions: Basal only (1), and Basal-Supplementary (4). The interpretations of these interactions are that:

- The Basal only type of program was more effective with below average students than the Basal-Linguistic/Phonics type, and less effective with above average students; it was less effective than the Basal-Individualized Reading type with above average students, and comparable in effectiveness to the Basal-Supplementary and Linguistic/Phonics only type;
- The Basal-Supplementary type of program was more effective than the Basal-Linguistic/Phonics type with below average students;

TABLE XXV

STATISTICAL DATA FOR EACH READING PROGRAM COMPARED, GRADE 4

Program	Variable	Students	Unadjusted Mean	Unadjusted SD	Adjusted Mean	Regression Slope+	Rank of:+ Unadj. M.	Rank of:+ Adj. M.
Basal		65				.12	3	4
	Pretest % Obj's Mastered		44.1	22.2				
	Posttest %		75.8	18.1	73.1	.23		
	Reading Level*		9.4	1.7				
	Pre-Post Diff:		31.7					
Basal-Linguistic/Phonics		281				.72	5	5
	Pretest %		36.1	13.9				
	Posttest %		70.6	17.2	72.4			
	Reading Level		8.6	2.0				
	Pre-Post Diff:		34.5					
Basal-Individualized Reading		258				.61	4	2
	Pretest %		41.6	15.7				
	Posttest %		76.0	17.6	75.3	.09		
	Reading Level		8.7	1.5				
	Pre-Post Diff:		34.4					
Basal-Supplementary		116				.26	2	3
	Pretest %		43.0	15.4				
	Posttest		76.8	13.1	74.7			
	Pre-Post Diff:		33.8					
Linguistics/Phonics		60				.31	1	1
	Pretest %		38.0	12.5				
	Posttest %		76.6	12.8	78.1	.14		
	Reading Level		8.3	2.1				
	Pre-Post Diff:		38.6					

*Reading Level is a coded variable, not the actual level.
+The numbers come from the logit transformation analyses; the rest of the data in the table are from untransformed analyses.

TABLE XXVI

COMPARISON OF READING PROGRAMS BY RECEIVING EXTRA HELP, GRADE 4*

Program	Extra Help	Number of Students	Pretest %	SD	Posttest %(Unadj)	SD	Posttest % (Adj)	Pre-Post Diff(Unadj)	Reading Level	SD	Slopes Pre	Slopes RL
Basal	Yes	30	41.5	14.8	75.8	16.2	74.9	34.3	8.9	1.6	.43	1.44
	No	35	46.4	27.0	75.7	19.9	71.5	28.3	9.7	1.7		
Basal-Linguistics/Phonics	Yes	103	39.3	10.7	73.7	15.2	73.1	34.4	9.3	1.8		
	No	178	34.2	15.2	68.8	18.0	72.0	34.6	8.2	2.0		
Basal-Individualized Reading	Yes	92	40.8	15.8	73.7	19.3	73.1	32.9	8.9	1.2		
	No	166	42.1	15.7	77.3	16.5	76.5	35.2	8.6	1.7		
Basal Supplementary	Yes	58	44.5	17.3	77.8	11.9	74.9	33.3	9.4	1.4		
	No	58	41.6	13.1	75.9	14.2	74.6	34.3	9.1	1.5		
Linguistic/Phonics	Yes	4	45.3	2.3	80.4	2.6	76.6	35.1	9.6	0.5		
	No	56	37.5	12.8	76.4	13.2	78.2	38.9	8.2	2.2		
Average		780 (Total)	39.8		74.2		74.2	34.4	8.8			

Note: Reading Level (RL) is a coded variable, not the actual reading level recorded by the teachers.
*Includes students in multi-level classes whose grade level averages 3.5-4.4.

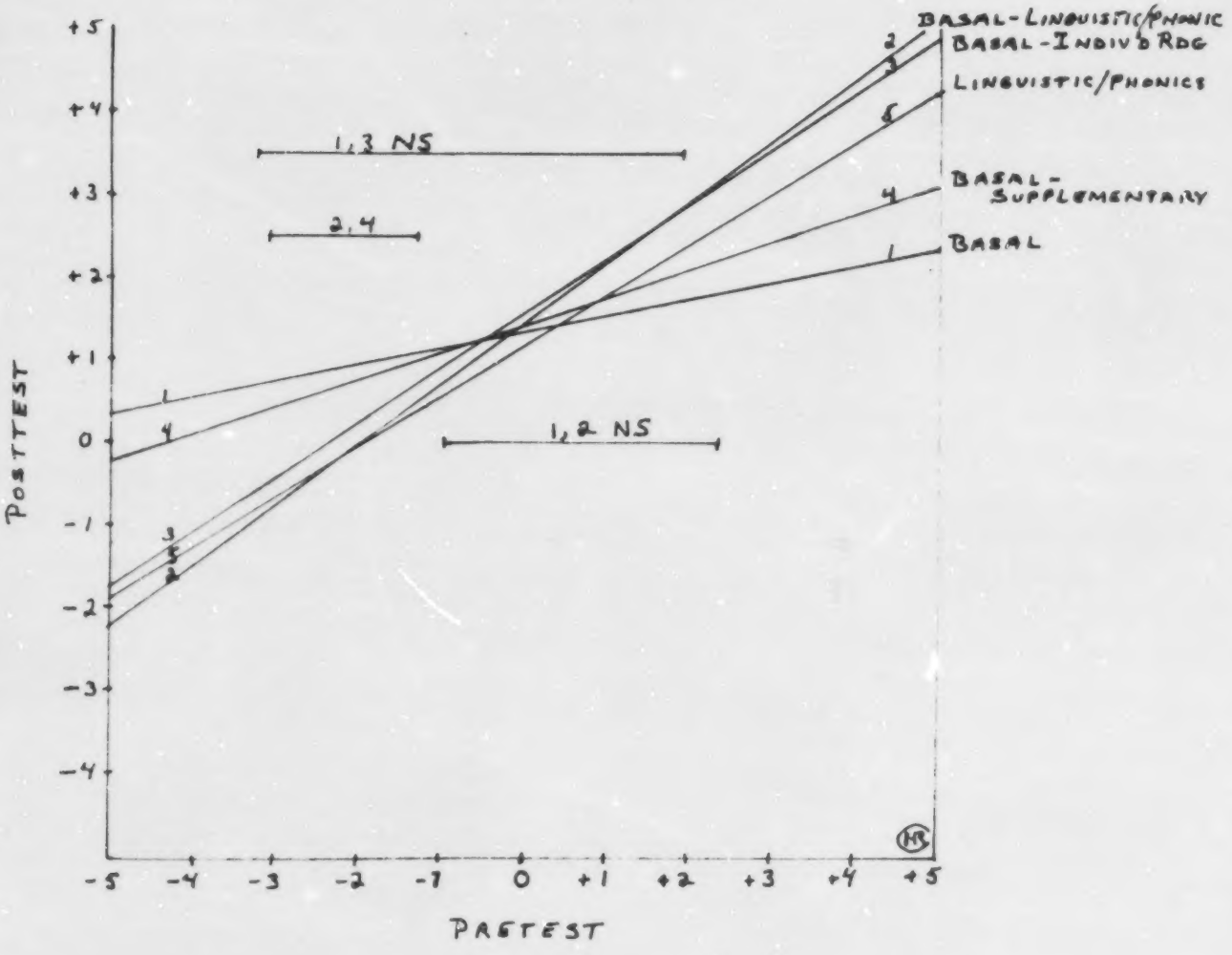
TABLE XXVII

COMPOSITION OF SAMPLES FOR ANALYSES OF READING PROGRAMS, GRADE 4

Program	Number of Classes	Number of Students	Average Number of Students Per Class	Minimum Number of Students Per Class	Maximum Number of Students Per Class
Basal	14	65	4.64	1	16
Basal-Linguistics/Phonics	36	81	7.81	1	17
Basal-Individualized Reading	36	258	7.17	1	22
Basal-Supplementary	25	116	4.64	1	23
Linguistics/Phonics	8	60	7.50	1	21
Total Number of Different Classes, Students	68	780			

Figure 3

Regions of Significant Difference Between Pairs of Reading Programs, Grade 4.



Note:

Because of the large number of interactions, bars are used to depict regions of significant differences or of non-significance (NS) between the regression slopes of reading programs. Numbers designate the pairs of programs to which the bars pertain.

· The remaining five pairs of programs were not significantly different; that is, they were comparable in performance with most students.

Over a two year period (1976-1978), there have been both consistencies and inconsistencies in performance relationships among these types of reading programs. Consistencies have been these:

The Basal only type has been less effective with above average students than the Basal-Individualized Reading and the Linguistic/Phonics only type; it has been comparable to the Basal-Supplementary type with all students.

· Linguistic/Phonics only has appeared to be more effective with above average students than: Basal only, Basal-Linguistic/Phonics, and Basal Supplementary types of programs; it has tended to be more effective with students in the mid-range of pretest than Basal-Individualized Reading.

· Basal-Linguistic/Phonics has been comparable in effectiveness with Basal-Individualized Reading and less effective with above average students than Linguistic/Phonics only.

· Basal-Individualized Reading has been less effective with average students than Linguistic/Phonics only, but more effective with above average students than Basal only; it has been comparable in effectiveness to Basal-Linguistic/Phonics and Basal-Supplementary with most students;

· Basal-Supplementary has been less effective than Linguistic/Phonics only for above average students; and it has been comparable in effectiveness to Basal only and Basal-Individualized Reading.

Inconsistencies in performance relationships over the two year period, 1976-78, have occurred between the following pairs of types of reading programs:

· Basal only and Basal-Linguistic/Phonics;

· Basal-Linguistic/Phonics and Basal-Supplementary.

The Linguistic/Phonics type of program was not included in the 1975-76 analyses. Over a three year period, 1975-78, there have been the following consistent patterns of performance relationships:

Basal only has tended to be less effective than a Basal-Individualized Reading type of program with above average students;

Basal-Linguistics/Phonics has been comparable in performance to a Basal-Individualized Reading type of program;

The Basal only type of program, as with the 3rd grade comparisons, either had inconsistent relationships over a two year period, or did not seem to have as good performance as some other types. There was a smaller size in 1977-78 than in 1976-77 (65 compared to 128). The Basal only group had the highest average pretest score of the groups compared in 1977-78 (see Table XXV); it had the next to lowest for 1976-77. The regression slope of the group was considerably flatter in 1977-78 than in 1976-77, as was also the case for Grade 3. Students comprising the group did not come from the same teachers each year (except for one of the 65 students in 1977-78). It is possible that sample selection factors have contributed to the results obtained. It is also possible that different teachers handled classes differently within a Basal only general approach to teaching reading from year to year. Both, of course, can be possible together.

The Linguistic/Phonics only approach has, for the past two years, involved the smallest samples of students. In 1977-78, the Linguistic/Phonics only group had the lowest pretest mean and average reading level of most frequentiy used material of any of the groups. In 1976-77, while average reading level was lowest, average pretest score was intermediate among the other groups. With one exception, the same teachers were not

involved both years; the one teacher accounted for 16 students (27%) of the group of 60 in 1977-78 and 13 of the students (28%) in the 1976-77 group. Despite these variations from year to year, the Linguistic/Phonics only group has maintained a fairly similar relationship in performance to other groups, doing relatively well especially for average to above average students. The regression slope for that type of program was a little flatter in 1977-78 than in 1976-77, which accounts for some changes from significance to non-significance in some comparisons of regions of significance. General performance relationships, however, tended to continue.

The results of the two-way ANCOVA were the same as for the '75-'76 and '76-'77 data. The effect of reading program was significant; the effect of receiving extra help in reading was not. There was not a significant interaction between the two variables. In effect, when type of program was taken into account, and when adjustment was made for pretest and level of most frequently used reading material, students who received extra help in reading seemed to gain at about the same rate as those who did not. The gap did not close, but neither did it tend to widen.

F. Comparisons of Reading Programs at the Fifth Grade Level

Five types of reading program were compared:

- Basal only
- Basal-Linguistic/Phonics
- Basal-Individualized Reading
- Basal-Supplementary
- Linguistic/Phonics only

Table XXVIII gives the descriptive statistics for each program. The one-way test of unadjusted means was not significant ($F_{4,621} = 1.813$, $p = .125$). The one-way test of adjusted posttest means was significant at

TABLE XXVIII

STATISTICAL DATA FOR EACH READING PROGRAM COMPARED, GRADE 5

Program	Variable	Number of Students	Unadjusted Mean	Unadjusted SD	Adjusted Mean	Regression Slope+	Rank of:+ Unadj. M.	Rank of:+ Adj. M.
Basal		38						
	Pretest % Obj's Mastered		53.2	21.4		.36		
	Posttest %		72.7	15.0	69.6		2	3
	Reading Level*		10.6	1.5		.16		
	Pre-Post Diff:		19.5					
Basal-Linguistic/Phonics		359						
	Pretest %		44.8	17.2		.55		
	Posttest %		73.9	15.5	74.5		1	1
	Reading Level		9.8	1.6		-.02		
	Pre-Post Diff:							
Basal-Individualized Reading		97						
	Pretest %		46.1	18.6		.59		
	Posttest %		72.8	14.7	72.9		3	2
	Reading Level		9.6	1.5		-.04		
	Pre-Post Diff:		26.7					
Basal-Supplementary		97						
	Pretest %		48.2	20.2		.42		
	Posttest %		70.7	15.3	69.9		5	4
	Reading Level		9.9	1.8		-.01		
	Pre-Post Diff:		22.5					
Linguistic/Phonic		35						
	Pretest %		50.6	19.6		.79		
	Posttest %		68.3	18.7	66.4		4	5
	Reading Level		10.0	1.3		.09		
	Pre-Post Diff:		17.7					

*Reading Level is a coded variable, not the actual level.

+These numbers come from analyses of logit transformations; the rest of the data in the table are from untransformed analyses.

the .05 level ($F_{4,619} = 5.120$, $p = .001$). Regression slopes were equal ($F_{8,611} = 1.399$, $p = .194$).

Descriptive statistics for the two-way analyses of covariance of groups are given in Table XXIX. In this analysis also there was a significant effect for reading programs ($F_{4,614} = 6.337$, $p = .000$), but not for receiving extra help ($F_{1,614} = 2.910$, $p = .089$). The interaction was significant ($F_{4,614} = 5.160$, $p = .000$). Table XXX shows the sample compositions for the programs compared.

There is no need for a figure showing the regression lines of the 5th grade programs. They are essentially parallel, as the results of the test of homogeneity of regression indicates. The information about effectiveness is contained in the separation between the lines, which is measured by the differences between means as long as linearity is assumed. From t-tests of differences between adjusted means it was found that the Basal-Linguistic/Phonics type of program was significantly more effective than Basal-Supplementary. All other pairs of program types were comparable in effectiveness (i.e., not significantly different).

Comparisons of the programs over a two year period, showed consistent trends in performance relationships:

· A Basal-Linguistic/Phonics type of reading program has tended to be more effective than Basal-Supplementary; it has been comparable to Basal-Individualized Reading, Basal only, and Linguistic/Phonics only;

· A Basal-Individualized Reading type of program has tended to be more effective than Basal only; it has been comparable to Basal-Supplementary, Basal-Linguistic/Phonics, and Linguistic/Phonics only;

A Basal only type of program has been comparable in effectiveness to: Basal-Linguistic/Phonics, Basal-Supplementary, and Linguistic/Phonics only

TABLE XXIX

COMPARISON OF READING PROGRAMS BY RECEIVING EXTRA HELP, GRADE 5*

Program	Extra Help	Number of Students	Pretest %	SD	Posttest % (Unadj)	SD	Posttest % (Adj)	Pre-Post Diff (Unadj)	Reading Level	SD	Slopes Pre	Slopes RL
Basal	Yes	18	59.1	20.9	73.4	15.4	68.0	14.3	10.7	0.6	.42	-.00
	No	20	47.8	21.0	72.0	14.9	71.4	24.2	10.6	2.1		
Basal-Linguistics/Phonics	Yes	175	49.5	16.1	77.8	13.6	76.5	28.3	10.3	1.1		
	No	184	40.4	17.2	70.1	16.2	72.6	29.7	9.3	1.8		
Basal-Individualized Reading	Yes	28	44.0	19.1	64.4	16.4	65.4	20.4	10.0	0.9		
	No	69	46.9	18.4	76.2	12.5	75.9	29.3	9.5	1.7		
Basal-Supplementary	Yes	55	46.7	18.1	70.8	12.5	70.6	24.1	10.3	0.8		
	No	42	50.1	22.7	70.7	18.6	69.1	20.6	9.5	2.6		
Linguistic/Phonics	Yes	12	38.0	14.3	57.2	16.3	60.7	19.2	9.0	1.6		
	No	23	57.2	19.0	74.1	17.5	69.5	16.9	10.5	0.8		
Averages		624 (Total)	46.3		72.8		73.0	26.5	9.9			

Note: Reading Level (RL) is a coded variable, not the actual reading level recorded by the teachers.
*Includes students in multi-level classes whose grade level averages 4.5-5.4.

-131-

TABLE XXX

COMPOSITION OF SAMPLES FOR ANALYSES OF READING PROGRAMS, GRADE 5

Program	Number of Classes	Number of Students	Average Number of Students Per Class	Minimum Number of Students Per Class	Maximum Number of Students Per Class
Basal	6	38	6.33	2	14
Basal-Linguistic/Phonics	37	359	9.70	2	20
Basal-Individualized Reading	16	97	6.06	1	16
Basal-Supplementary	12	97	8.08	2	17
Linguistic/Phonics	4	35	8.75	1	19
Total Number of Different Classes, Students	59	626			

type; it has been less effective than Basal-Individualized Reading;

• A Basal-Supplementary type has been comparable to: Basal only, Basal-Individualized Reading, Basal-Linguistic/Phonics, and a Linguistic/Phonics only type of reading program.

The trends in rank orders for the two years show a Basal-Linguistic/Phonics type of program to be the most effective of those compared both years, and a Linguistic/Phonics only type to be the least effective. The former has had the largest sample size of students each year; the latter has had the smallest. Both years the Basal-Linguistic/Phonics group had among the lowest average pretest scores (it was the lowest for 1977-78), while each year the Linguistic/Phonics only group had the highest or next to highest average pretest score of the groups compared. Thus, the relative effectiveness of two types of programs may be a consequence in part of selection bias.

Fifty-seven percent (20 out of 35) of the students comprising the Linguistics/Phonics only group in 1977-78 came from one teacher's class. That was the same teacher from whose class 59% (19 out of 32) of the Linguistic/Phonics only group came in 1976-77. One teacher thus had a particularly strong influence on the net performance associated with the Linguistic/Phonics only type of program both years. In the Basal-Linguistic/Phonics group, ten of the same teachers were involved both years (out of 37 in 1977-78, and out of 32 in 1976-77). Those 10 teachers contributed 25% of the sample size of students (91 out of 359) in 1977-78, and 31% (98 out of 254) in 1976-77. The performance characteristics associated with the Basal-Linguistic/Phonics type of program were thus less subject to the influence of a very few individual teachers than the Linguistic/Phonics only type. The Basal only sample, which was small in

1977-78, did not involve any teachers contributing to the Basal only sample in 1976-77.

It is concluded here that while most of the program relationships over the two year period are probably reasonably representative, there is reason to regard the Linguistic/Phonics only performance as non-representative of a variety of classes and groups of students. It is thus inappropriate to conclude that it is the least effective in relative performance until larger and more generalizable samples confirm that it is.

Both in 1977-78 and 1976-77 there was a significant interaction between type of reading program and receiving extra help. Thus, both years, the effect of receiving extra help has depended on the type of program.

G. Comparisons of Reading Programs at the Sixth Grade Level

Four types of reading program were compared:

- Basal only materials
- Basal-Linguistics/Phonics materials
- Basal-Individualized Reading materials
- Basal-Supplementary materials

Table XXXI summarizes descriptive statistics for each program. The test of differences of unadjusted means for the one-way analysis of variance was not significant, nor was the one-way ANCOVA test of adjusted means ($F_{3,301} = .453$, $p = .716$; $F_{3,299} = 1.969$, $p = .119$). Regression slopes were ($F_{6,293} = .588$, $p = .740$).

Table XXXII gives descriptive statistics for the two-way ANCOVA. The variable, reading program, was not significant $F_{3,295} = 1.575$, $p = .195$); the variable, extra help, also was not $F_{1,295} = .516$, $p = .473$). The interaction was not significant ($F_{3,295} = .586$, $p = .624$).

TABLE XXXI

STATISTICAL DATA FOR EACH READING PROGRAM COMPARED, GRADE 6

Program	Variable	Number of Students	Unadjusted Mean	Unadjusted SD	Adjusted Mean	Regression Slope+	Rank of:+ Unadj. M.	Rank of:+ Adj. M.
Basal		35						
	Pretest % Obj's Mastered		59.1	18.2		.53		
	Posttest %		76.8	13.5	72.6		1	4
	Reading Level*		11.4	0.9		-.03		
Basal-Linguistic/Phonics		121						
	Pretest %		53.7	21.7		.44		
	Posttest %		72.9	16.0	71.9		3	3
	Reading Level		10.5	1.8		.10		
	Pre-Post Diff:		19.2					
Basal-Individualized Reading		51						
	Pretest %		48.9	24.1		.37		
	Posttest %		71.5	17.8	72.8		4	2
	Reading Level		10.5	1.8		.10		
	Pre-Post Diff:		22.6					
Basal-Supplementary		98						
	Pretest %	45.9	45.9	19.9		.25		
	Posttest %		75.0	16.6	77.0		2	1
	Reading Level		10.1	1.9		.08		
	Pre-Post Diff:		29.1					

*Reading Level is a coded variable, not the actual level.
+These numbers are from the analyses of logit transformations; the rest of the data in the table are from untransformed analyses.

TABLE XXXII

COMPARISON OF READING PROGRAMS BY RECEIVING EXTRA HELP, GRADE 6*

Program	Extra Help	Number of Students	Pretest %	SD	Posttest %(Unadj)	SD	Posttest % (Adj)	Pre-Post Diff(Unadj)	Reading Level	SD	Slopes Pre	Slopes RL
Basal	Yes	17	54.2	18.0	78.6	12.8	75.9	24.4	11.4	0.6	.32	1.52
	No	18	63.7	17.7	75.2	14.2	69.6	11.5	11.4	1.1		
Basal-Linguistic/ Phonics	Yes	72	59.5	19.6	75.5	14.5	71.8	16.0	11.0	1.3		
	No	49	45.2	21.9	69.1	17.4	72.0	23.9	9.7	2.1		
Basal-Individualized Reading	Yes	25	54.1	23.4	74.6	14.3	72.9	20.5	10.8	1.9		
	No	26	43.9	24.1	68.5	20.4	72.6	24.6	9.2	2.5		
Basal-Supplementary	Yes	62	48.3	20.2	77.8	13.8	78.1	29.5	10.7	1.4		
	No	36	41.8	19.0	70.0	19.8	75.1	28.2	9.0	2.2		
Average		305 (Total)	51.0		73.8		73.8	22.8	10.4			

Note: Reading Level (RL) is a coded variable, not the actual reading level recorded by the teachers.
*Includes students in multi-level classes whose grade level averages 5.5-6.4.

Table XXXIII gives the sample compositions for programs in these analyses.

The main finding from the one-way ANCOVA is that there were no significant differences in performance among the four types of reading programs compared.

For the 4 types of programs compared both in 1976-77 and 1977-78, there have been both consistencies and inconsistencies of performance relationships. The consistencies were that:

- Basal only has been comparable in effectiveness to Basal-Linguistics/Phonics and to Basal-Individualized Reading,
- Basal-Linguistic/Phonics has been comparable in effectiveness to Basal-Individualized Reading and to Basal only;
- Basal-Individualized Reading has been comparable to Basal only and Basal-Linguistic/Phonics.

These three types, in effect, have all been comparable in performance over the two year period.

Inconsistent relationships in performance have been observed between:

Basal-Supplementary and: Basal only, Basal-Linguistic/Phonics, Basal-Individualized Reading.

The immediate reason for the inconsistencies is that in 1976-77 the Basal-Supplementary group had a markedly flatter regression line than the other types of programs, with resultant regions of significant difference among Basal-Supplementary and the other types. That was not the case in 1977-78, when regressions were homogeneous (the lines essentially parallel with each other).

The same consistent relationships described above occurred with the 1975-76 comparisons also, except that Basal-Linguistic/Phonics was

TABLE XXXIII

COMPOSITION OF SAMPLES FOR ANALYSES OF READING PROGRAMS, GRADE 6

Programs	Number of Classes	Number of Students	Average Number of Students Per Class	Minimum Number of Students Per Class	Maximum Number of Students Per Class
Basal	5	35	7.00	1	13
Basal-Linguistic/Phonics	27	121	4.48	1	19
Basal-Individualized Reading	14	51	3.64	1	11
Basal-Supplementary	21	98	4.67	1	17
Total Number of Different Classes, Students	51	305			

significantly more effective than Basal-Individualized Reading for students scoring below average on pretest. The regression line for Basal-Supplementary was the flattest of all groups, though not as markedly so as in 1976-77.

Six of the same teachers contributed to the Basal-Supplementary samples of students in 1976-77 and 1977-78. That was 26% of the number of different teachers in 1976-77 and 29% in 1977-78. Twelve percent of the 134 students in the Basal-Supplementary sample came from those teachers in 1976-77; 34% of the 98 students came from them in 1977-78. The Basal-Supplementary group had the lowest pretest average in 1977-78; it was slightly above the average of all groups compared in 1976-77.

These data suggest that the Basal-Supplementary samples were influenced by some of the same teachers both years, but they were not a majority of the teachers involved. The low standing with respect to average pretest score in 1977-78 suggests the possibility of some selection bias as an influence on performance. However, neither of those factors clearly accounts for the change in regression slopes (of posttest regressed on pretest) between the two years.

It would appear that, except for the Basal-Supplementary type of program, the other types compared both years can be regarded as comparable to each other. It will be necessary to see if the 1978-79 data clarify a trend for the Basal-Supplementary type.

In the two-way ANCOVAs of 1976-77 and 1977-78, neither main effects (type of program, extra help) nor interactions have been significant. The groups receiving extra help in reading have gained comparably to those not receiving it.

H. General

It is of interest to note that average reading level of materials increased at about the same rate, in terms of actual grade levels, at the primary grades and at the intermediate ones. The average grade levels of reading materials, by grade, are:

Grade 1: between Primer and 1^1 (grade 1, first semester)

Grade 2: slightly over 1^2 (grade 1, second semester)

Grade 3: slightly over 2^2

Grade 4: about halfway between 3^1 and 3^2

Grade 5: about grade 4

Grade 6: about halfway between 4 and 5

These, of course, are grade levels according to teachers' ratings of the materials used with the students involved in the analyses here. The average grade levels of the reading materials in the '77-'78 data were about the same as in '76-'77.

Before summarizing the results of the two years of evaluations of selected types of reading programs, it is appropriate to comment on the decision criteria employed, and on the implications of the findings for subsequent analyses (e.g., analyses that will include data from school year 1978-79).

All statistical tests have been tests of the null hypothesis, with the .05 level used as the overall test criterion. Any difference observed with a p-value of .050 or less was labelled significant; any difference with a p-value of greater than .050 was labelled not significant. As a general principle in this report, this has been a protected level in that if an overall test did not indicate significance, no sub-test was interpreted regardless of its p-value. If an overall test of differences was significant at the .05 level, then tests of differences of pairs within

set used p = .05/n (where n = the number of tests to be made) as the critical value for any one test. That is a Bonferroni procedure, intended to protect a family wise Type I error rate. For example, in Grade 5 the overall test of difference among five programs was significant at the .05 level, and the test of homogeneity of regression was not significant. Thus, each test of differences of the n (n - 1)/2 = 10 pairs used a significance level of .005 (= .05/10). This procedure was also used in the Neyman-Johnson tests of regions of significant differences when it was found that there was not homogeneity of regression. Thus, significance had the same criterion in either case. It is a very conservative procedure.

The null hypothesis is often taken as the hypothesis that the observed value is probably not really different from zero, or that the observed difference could have occurred by chance 19 out of 20 times. The validity of tests of the null hypothesis rests on the assumption that samples were in fact drawn from the same population. Thus, the p-value of the statistical test is not the probability that the null hypothesis is true, given the difference observed. It is the probability of the difference occurring if the null hypothesis is true (cf. Cronbach and Snow, 1977; Carver, 1978). The probability that the null hypothesis is true, given the observed difference, is a problem for Bayesian statistics, as Cronbach and Snow point out, not for classical statistics.

In much of behavioral research, and certainly in the case of non-random samples such as those employed here, there is no way to know if the null hypothesis is true. A strict conclusion, therefore, is that the use of significance tests such as those employed here is not only irrelevant but wrong. While that conclusion is technically and theoretically correct, however, it is unsatisfactory since it gives no indication of how

to proceed in decision-making in the alternative. Meanwhile, school personnel not only can but do make important decisions affecting students' lives and the utilization of resources on all manner of bases, some of which are at least as challengeable from technical and theoretical points of view as the decision process used here.

The position taken by the Evaluation System is to assume that the null hypothesis is true for samples unless there is reason to think otherwise,* and to use statistical criteria as crude pointers which essentially indicate whether or not there is a reason to take a distinction seriously. But it also insists on repeatability before suggestions or recommendations about relative effectiveness are made. Thus, the reader has seen in the preceding pages that detailed attention has been paid not just to the significance tests, but also to whether or not patterns of results have been consistent over two consecutive years. Analyses of '78-'79 data will of course give yet further evidence bearing on conclusions, but in the meantime, those consistencies that have emerged over a two to three year period are considered sufficient to warrant making suggestions for teachers' and administrators' consideration. While the school system must do its best to avoid and ignore tenuous findings, neither can it wait until all the data are in, since all the data will never be in. Thus, the Evaluation System's approach is to try to balance theoretical purity with practicality by means of a clearly describable, conservative procedure and set of decision rules.

As a final point, the results of the '75-'76, '76-'77, and '77-'78 analyses of types of reading programs strongly argue for the continuation

---

*Possible sampling and selection biases have been discussed in the presentation of results for each grade level.

and expansion of the Evaluation System. Sufficient evidence has been found over the three years, taken together, of differential and comparable effectiveness associated with different categories or types of programs to warrant continuing with the methodology, especially as the CBC becomes widely used. For types of programs that have been found to be particularly effective or ineffective, it would be appropriate to apply the same ANCOVA analysis to the specific reading materials employed. It would also be appropriate to analyze the bases of interactions more precisely, examining variables other than just pretest. It would, in addition, be appropriate to examine effects with different types of students in more detail. For example, it is appropriate now to include in analyses students who were administered different levels of the PRT pre and post, and analyze the relationship of certain classroom level variables to reading program performance.* Finally, it is appropriate to start examining longitudinal results with the same students, as the Evaluation System is designed to do. Since there is now a basis for confidence in the ability of the Evaluation System to identify relative effectiveness and factors associated with it, there is increased justification for continuation and expansion. Expansion is important not for its own sake, but to produce samples adequate to support analyses of increasingly specifically defined programs and conditions of instruction, and to reduce the Type II error rate now associated with the limited sample sizes.

*Both such analyses have also been done on exploratory bases using the 1976-77 data. Again, results have been encouraging and warrant further such analyses.

I. Summary

The evaluation of types of reading programs has involved comparing the adjusted posttest reading achievement scores (percentages of reading objectives mastered) of samples of students whose teachers used the same types of reading materials with those students. Students at each grade level were considered to have the same type of reading program if their most frequently used and next most frequently used reading materials were of the same types. These are illustrations of types of reading programs:

- Basal-Language Experience: a basal reader and a language experience material (for example: a Bank Street Reader supplemented with Experience Charts);
- Basal-Linguistic/Phonics: a basal reader and a linguistic/phonics material (for example: a Harper and Row Basic Reader supplemented by Lyons and Carnahan's Phonics We Use)
- Basal-Individualized Reading: a basal reader supplemented with an individualized reading material (for example, a Holt, Rinehart, Winston Basic Reading System material in conjunction with SRA Reading Lab material);
- Linguistic/Phonics-Supplementary: a linguistic/phonics material with a material classified as Supplementary (for example, Phono Visual supplemented with the D. C. Schools' Reading Comprehension Skills Exercises).

The Evaluation System's Student Reading Program Form in Appendix B lists a variety of materials classified under each main category or type. Table XXXIV shows by grade level the types of programs evaluated in 1975-76, 1976-77, and 1977-78. For technical and professional reasons, conclusions can be drawn about differences in effectiveness, and recommendations or suggestions made, only when particular patterns of results continue for at least two consecutive years, with the pattern being statistically significant at least one of those years. When patterns change, data from

TABLE XXXIV

TYPES OF READING PROGRAMS EVALUATED, 1975-76, 1976-77, and 1977-78, BY GRADE*

Type of Reading Program	Grade 1	2	3	4	5	6
1. Basal only	5,6	5,6	5,6,7	5,6,7	5,6,7	5,6,7
2. Basal-Linguistic/ Phonics	5,6,7	5,6,7	5,6,7	5,6,7	5,6,7	5,6,7
3. Basal-Programmed	5,6,7	5,6,7	5,7			
4. Basal-Language Experience	5,6,7	6,7	6,7		5,6	
5. Basal-Individualized Reading		6,7	5,6,7	5,6,7	5,6,7	5,6,7
6. Basal-Supplementary	5,7	5,6,7	6,7	5,6,7	5,6,7	5,6,7
7. Basal-Special	7					
8. Linguistic/Phonics Only	6,7	5,6,7	6,7	6,7	6,7	5,6
9. Linguistic/Phonics-Basal	5	6,7				
10. Ling/Phon-Indiv'd Reading				6	5	
11. Linguistic/Phonics-Supplementary	6					6
12. Supplementary Only						6
13. Programmed-Basal		5				

*Numbers indicate year the type of program was evaluated:
5 = 1975-76
6 = 1976-77
7 = 1977-78

the subsequent year or years will be needed to ascertain probable trends. By those criteria, the following are the conclusions to be drawn to date from evaluations of the 1975-76/1976-77/1977-78 data samples at each grade level. They pertain only to those types of reading programs with sufficient sample sizes (30 students or more) to be included in the evaluations at least two consecutive years (1976-77, 1977-78).

1. Grade One

a. Four types of programs have been evaluated both in 1976-77 and in 1977-78: Basal-Linguistic/Phonics, Basal-Programmed/Structured, Basal-Language Experience, and Linguistic/Phonics only.

b. There have been no significant differences in effectiveness on the average associated with Basal-Linguistic/Phonics and Basal-Programmed/Structured when they were compared with each other. Indeed, this result has held over a three year period.

The evidence thus indicates that teachers can use whichever type they believe is best for them and their students with the understanding that on the average these two types have been found to be comparable in effectiveness. Decision-makers should understand, however, that the results for the Basal-Programmed/Structured approach have been based on relatively small samples of students from a small number of teachers who may have been especially trained in the approach. Thus, the training and experience of the teacher should be given particular consideration in deciding among approaches.

c. The results for a Basal-Language Experience type of program over two years have not been consistent relative to the other types of programs. Students in that type of program have, however, consistently shown the

largest absolute gains both years, compared to students in the other programs. As with the Basal-Programmed/Structured approach, the Basal-Language Experience results may have shown the influence of the work of a small number of teachers. Thus, again, particular attention should be given to the familiarity of teachers with the approach when making a decision about employing it.

d. Results for the Linguistic/Phonics only type of reading program have been mixed over the period of 1976 to 1978. Based on adjusted post-test means, it ranked next to lowest in 1976-77, and highest in 1977-78. But it also had the lowest average pretest scores in 1976-77, and above average scores in 1977-78. Furthermore, sample sizes have been small compared to samples in the other programs compared. Data from further samples are needed before conclusions should be drawn about the relative effectiveness of this type of program.

2. Grade Two

a. The types of programs evaluated for two consecutive years were: Basal-Linguistic/Phonics; Basal-Programmed/Structured; Basal-Language Experience; Basal Individualized Reading; Basal Supplementary; Linguistic/Phonics only; and Linguistic/Phonics-Basal.

b. The evidence over two years suggests that Linguistic/Phonics-Basal type of program has tended to be more effective than a Basal-Programmed/Structured, a Basal-Individualized Reading and a Basal-Language Experience type, especially for students who score average or above on the PRT in the fall; and more effective than a Basal-Supplementary type, especially for students who score average or below in the fall. Again, however, it is important to keep in mind that the results for Linguistic/

Phonics-Basal may be a combination of teacher and program effects. Thus, teacher training and experience should be particularly considered in decision-making about that type of program.

c. Over the periods 1976-77 and 1977-78, performance relationships have been inconsistent between the following types of programs:

1) Basal-Linguistic/Phonics and: Basal-Language Experience, Basal-Individualized Reading, and Basal Supplementary;

2) Basal-Programmed/Structured and Linguistic/Phonics only;

3) Basal-Language Experience and: Basal-Linguistic/Phonics, Linguistic/Phonics only;

4) Basal-Individualized Reading and: Basal-Linguistic/Phonics, Linguistic/Phonics only;

5) Basal-Supplementary and Linguistic/Phonics only;

6) Linguistic/Phonics only and: Basal-Programmed/Structured, Basal-Language Experience, Basal-Individualized Reading, Basal-Supplementary.

Thus, the Evaluation System can not yet provide conclusions or statements about the relative effectiveness of those alternatives.

d. The evidence over two consecutive years indicates reason to believe that the following groups of program types tend to produce comparable results, as measured by adjusted percentage of objectives mastered in the spring on the PRT:

1) Basal-Linguistic/Phonics and: Basal-Programmed/Structured, Linguistic/Phonics only, Linguistic/Phonics-Basal;

2) Basal-Programmed/Structured and: Basal-Linguistic/Phonics, Basal-Language Experience, Basal-Individualized Reading, Basal-Supplementary materials;

3) Basal-Language Experience and: Basal-Programmed/ Structured, Basal-Individualized Reading, Basal Supplementary

4) Basal-Individualized Reading and: Basal-Programmed/ Structured, Basal-Language Experience, Basal-Supplementary;

5) Basal-Supplementary and: Basal-Programmed/Structured, Basal-Language Experience, Basal-Individualized Reading;

6) Linguistic/Phonics only and: Basal-Linguistic/Phonics, Linguistic/Phonics-Basal;

7) Linguistic/Phonics-Basal and: Basal-Linguistic/Phonics, Linguistic/Phonics only.

It should be noted that samples of student in both years have been such that for Basal-Programmed/Structured, and Linguistic/Phonics only, results may have been particularly influenced by a teacher effect as well as a program effect.

3. Grade Three

a. Six types of programs had adequate samples sizes to be included in the analyses both years: Basal only, Basal-Linguistic/Phonics, Basal-Language Experience, Basal-Individualized Reading, Basal-Supplementary, and Linguistic/Phonics only.

b. Over the two year period, results have indicated that for students who scored above average on pretest, a Linguistic/Phonics only type of program has been more effective than a Basal only, Basal-Linguistic/Phonic, Basal-Language Experience, Basal-Individualized Reading, and Basal-Supplementary type. There have also been indications that for students who score very low on pretest, a Linguistic/Phonics only type of program may have been less effective than those listed. The Linguistic/Phonics only samples each year were small and had low pretest averages, compared to the

other groups. Their results may also have been more influenced by teacher effects as well as program effects than some of the other types of programs compared.

c. Results have indicated that a Basal-Individualized Reading type of program has been associated with greater effectiveness than a Basal-Linguistic/Phonics and a Basal-Language Experience type with students who score below average on pretest; it has been comparable in performance with a Basal only and Basal-Language Experience type for students throughout the pretest range.

d. Over the two-year period the following have been comparable in effectiveness: Basal-Linguistic/Phonics, Basal-Supplementary, and Basal-Language Experience.

e. A Basal only type of program has been comparable in effectiveness to a Basal-Individualized Reading type.

f. Based on two years of comparisons, the performance relationships of the following have been inconsistent:

1) Basal only and: Basal-Linguistic/Phonics, Basal-Language Experience, Basal-Supplementary;

2) Basal-Individualized Reading and Basal-Supplementary.

The Basal only groups for the two years may, like the Linguistic/Phonics only groups, have particularly shown the influence of a small number of teachers, as well as the effects of the type of program. The samples also had different average pretest characteristics each year, being somewhat below average in '76-'77 and next to highest in '77-'78. Therefore, further samples of students with more generalizable characteristics for that type of program will be needed to establish consistent relationships with other types. Similar considerations apply to the Basal-Individualized Reading type.

4. Grade Four

a. Five types of programs have been compared both years: Basal only, Basal-Linguistic/Phonics, Basal-Individualized Reading, Basal-Supplementary, and Linguistic/Phonics only.

b. Both in 1976-77 and 1977-78, a Basal only type of program has been less effective with above average students than a Basal-Individualized Reading and a Linguistic/Phonics only type; it has been comparable in effectiveness to a Basal-Supplementary type with most students.

c. The evidence suggests that a Linguistic/Phonics only type of program is more effective with above average students than a Basal only, Basal-Linguistic/Phonics, or a Basal-Supplementary type; it has tended to be more effective with average students than a Basal-Individualized Reading type. There may have been some teacher effect as well as program effect in the Linguistic/Phonics only sample results for the two years, but performance relationships of that type of program to others have been consistent.

d. A Basal-Linguistic/Phonics type of program has been comparable in effectiveness to a Basal-Individualized Reading type; it has been less effective with above average students than a Linguistic/Phonics only type.

e. A Basal-Individualized Reading type of reading program has been comparable to a Basal-Linguistic/Phonics and a Basal-Supplementary type with most students. Indeed, those relationships have occurred over three years as well as the past two.

f. Over a two-year period, the performance relationship have not been consistent between the following:

1) Basal only and Basal-Linguistic/Phonics;

2) Basal-Linguistic/Phonics and Basal-Supplementary.

5. Grade Five

a. Five types of programs have been compared both years: Basal only, Basal-Linguistic/Phonics, Basal-Individualized Reading, Basal-Supplementary, and Linguistic/Phonics only.

b. With two exceptions, the five types have been comparable to each other in effectiveness over a two year period: Basal only, Basal-Linguistic/Phonics, Basal-Individualized Reading, Basal-Supplementary, and Linguistic/Phonics only. The exception has been that a Basal-Individualized Reading type of reading program has given evidence of being more effective than a Basal only type.

c. A Linguistic/Phonics only type is the most likely of the four programs to have been influenced by the work of the same teacher each year. Thus, the performance results for this type of program are the most likely of the four to reflect particular teacher effects as well as program effects.

6. Grade Six

a. Four types of programs have been evaluated both years: Basal only, Basal-Linguistic/Phonics, Basal-Individualized Reading, and Basal-Supplementary.

b. The following types of programs have been comparable to each other in performance effectiveness for three years: Basal only, Basal-Linguistic/Phonics, and Basal-Individualized Reading.

c. Performance relationships have been inconsistent over a three-year period between a Basal-Supplementary type of program and the other three types listed in b. The Basal-Supplementary type of program results may particularly have been a combination of teacher and program effects, as well as of the effects of particular types of students involved. Further, more generalized samples are needed to clarify performance characteristics of the Basal-Supplementary type of program at this grade level.

7. The results of analyses of the 1976-77 and 1977-78 data suggest that students who have received extra help in reading have gained at the same rate as those who have not. There is indication that at least at some grade levels the effectiveness of extra help may depend on the main type of reading program that the student receives. Students who have received extra help have typically, but not always, been those who scored below those who did not on the fall PRT. Thus, while gaps between those two categories of students have generally not closed at posttest, neither have they widened.

8. Decision-makers should bear in mind that all conclusions about program effectiveness have been based on analyses of samples of students who were given the same level of PRT in the fall and in the spring. Future evaluations will include students who were given different pre-post PRT test levels.* Future evaluations will also take into account more specific characteristics of reading programs than just the general types of materials, reading level of materials, and pretest scores of students. At this stage in the development of the Evaluation System, evaluations have been limited simply to comparisons of major types of reading programs that are the most prevalent in the D.C. Schools. It is assumed that different types of programs imply different theories and methods of reading instruction. The assumption, however, has not yet been tested empirically by the Evaluation System. Finally, no distinction has been made within program samples at a given grade level between students who received different levels of the PRT.

---

*Preliminary analyses have been made which included such samples. Some conclusions reported here would probably have been different had they been included. Pending a validated latent trait model, however, the students involved in the present analyses are the majority of students in the DCPS.

Since gain is correlated negatively with test level, different mixes of levels within programs could affect comparative performance outcomes.* Future analyses will continue to examine this potential source of bias.

9. In making decisions about choices of reading programs, decision-makers should therefore seriously consider the findings to date but not feel compelled to act upon them mechanically in all cases. As noted throughout, the findings are based on average outcomes. Furthermore, the decision-maker needs to keep in mind that statements about the effectiveness of types of reading programs mean average effectiveness associated with them, not necessarily caused by them. Other data considered pertinent and important in making program choices should also be taken into consideration.

---

* However, exploratory analyses using 1976-77 data have led to the conclusion that they do not.

VII CONCLUSIONS

Data from the 34 elementary schools in the Evaluation System were analyzed separately for the school years 1975-76, 1976-77, 1977-78. As with the 1976-77 report, the 1977-78 findings have been presented in a two-part report: The summary report and the technical report. From these findings main conclusions were drawn. Where appropriate, the findings from the 1976-77 report have been incorporated in this report.

A. Reading Achievement and Testing

1. In 1977-78, the average percentage of total possible reading objectives mastered in the fall on the Prescriptive Reading Test (PRT) by 489 classes of students in Grades 1-6 was 48%.* The average percent of total possible reading objectives mastered by the same classes and students in the spring was 76%. Thus, during the year, classes on the average showed an increase of 28% in the mastery of the total possible objectives.

2. The average percentage of objectives mastered at pretest in 1977-78 was markedly lower than for 1976-77 and for 1975-76. The average percentage of objectives mastered at posttest, however, was only slightly higher (1-2%) than for those two years. Therefore, one should not conclude that the average gain of 28% in 1977-78 was twice as good as the average gains of 14-15% in the two prior years. It is not yet possible to conclude why the average pretest scores were so much lower in 1977-78 than in the previous years. For the two previous years there had been no significant difference in average percentage of objectives mastered on the pretests.

3. In 1977-78, the average pretest percentage of reading objectives mastered was higher for Grades 1-3 than for Grades 4-6: 51% compared to

---

*All conclusions for all three years are based on analyses of data from students who received the same level of PRT in the fall and in the spring. They constitute about 70% of all students in the 34 schools.

44%. The average posttest percentage of objectives mastered was also higher for the primary than for the intermediate grades: 80% compared to 72%. Because of those differences, the average percentage of objectives gained for the two groups of grades was very similar: 29% compared to 27%. In the two previous years, classes in the primary grades had a markedly larger average gain than the classes in the intermediate grades: 19% compared to about 10%. But those differences were based on similar average pretest scores. Thus, it would not be defensible to draw conclusions about changes in the gains of the primary grades in relation to the intermediate grades.

4. For 1977-78, results of all six grades show that the greatest pre-post gain in reading scores occur in classes which:

a) have low levels of teacher turnover;

b) have low teacher absence rates;

c) have lower percentages of students receiving free lunch;

d) have fewer pupils needing extra reading help and relatively large numbers receiving extra reading instruction.

5. In 1977-78, homogeneity of grouping (low test level count) and greater amounts of self-initiated reading, particularly in the primary grades, are predictive of higher posttest performance scores.

6. For 489 classes, Grades 1-6, teachers administered an average of 2.1 different levels of the PRT per class in the fall and spring testing in 1977-78. That compares with 2.0 in 1976-77, and 2.2 in 1975-76. These are minimum averages, since in all three cases the averages are based only on students who took the same level of test in the fall and spring.

B. Interest in Reading

As has been reported previously, higher interest in reading, on a class average basis, is predictive of increased reading achievement. It

is therefore important to be concerned with those factors that have a high correlation with reading interest. Results of regression analysis for the 1977-78 year show that class average interest in reading is significantly greater in classes which:

a) have higher reading pretest scores. This was also true for 1975-76 and 1976-77.

b) have lower class absence rates, particularly in the intermediate grade level. This was also true for the two prior years.

c) have a lower percentage of male students.

d) have low teacher turnover and teacher absence rates especially at the primary grade level.

e) have a lower percentage of students receiving free lunch.

f) have a lower need for extra help in reading. True also for two prior years.

g) have less variability in the levels of reading material used.

h) are self contained, as opposed to open space, particularly at the intermediate grade level.

C. Evaluations of Selected Types of Reading Programs

1. Commonly used types of reading program at each grade have been compared for two, and in some cases, three consecutive years (1975-76, 1976-77, and 1977-78). Using the results of these analyses it has been possible to suggest guidelines to be considered by decision-makers in selecting types of reading programs. The guidelines suggested for each grade level are listed under "Findings" in Section VI of this Summary Report.

2. The guidelines, which are based on the application of a proposed set of decision rules for interpreting the results of evaluations for two consecutive years, consist of three kinds of statements:

a) an identification of types of reading programs which have consistently appeared to be more, or less, effective than other types for specified types of students;

b) an identification of types of reading programs which have consistently appeared to be comparable in effectiveness for specified types of students;

c) an identification of types of programs about which the evidence is still unclear because of marked changes in relative effectiveness from one year to the next (further evaluations are needed in subsequent years to clarify the effectiveness relationships of these types of programs).

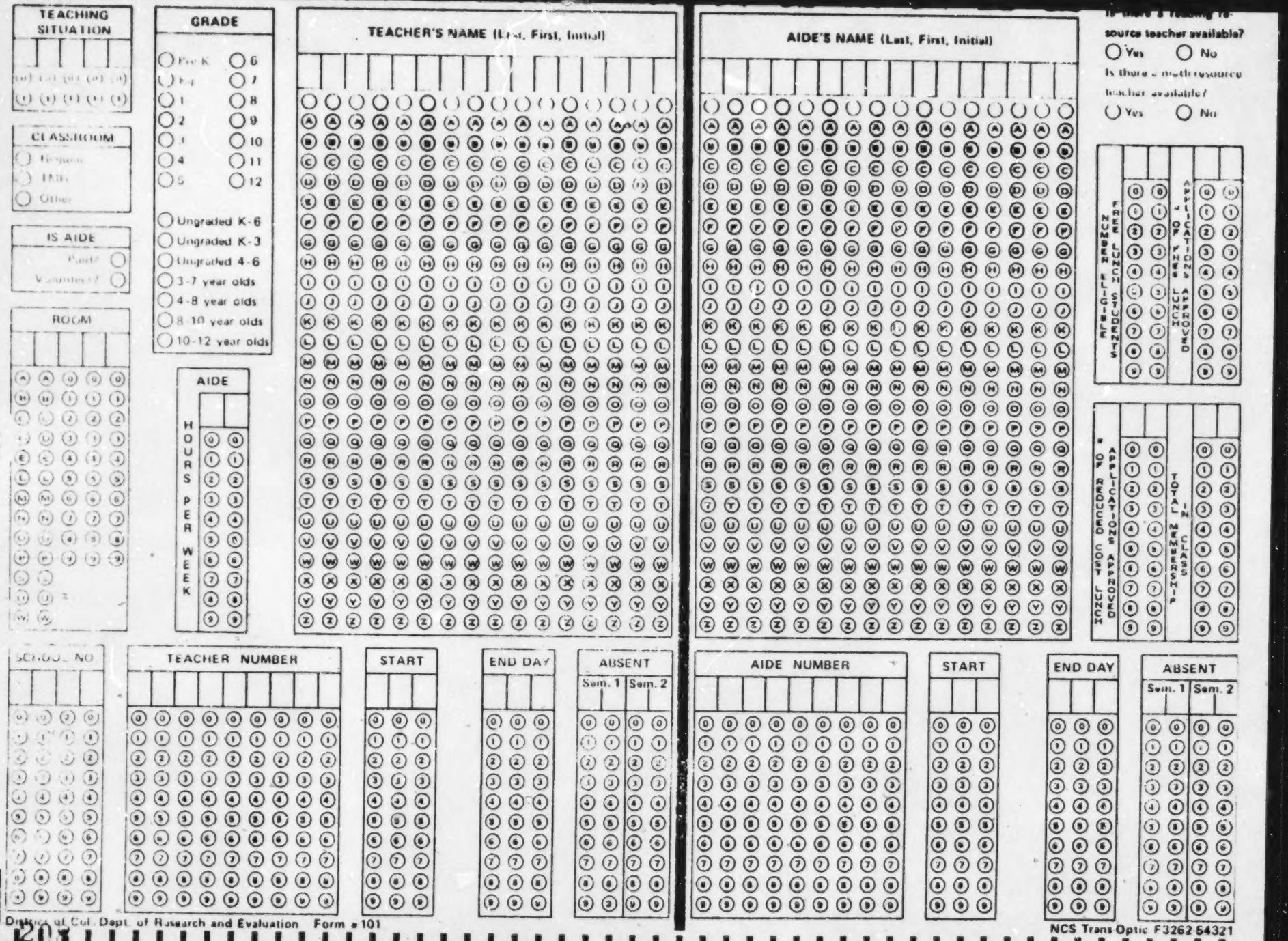
3. All guidelines have come from the analysis of pre and posttest PRT scores of students who were administered the same level of test in the fall and in the spring. They have come from measures of effectiveness consistently associated with different types of reading programs but not necessarily caused by them. Future evaluations of effectiveness will take into consideration students who were administered different levels of the PRT in the fall and in the spring. They will also take into consideration additional factors, other than pretest score and the grade level of the reading materials used, that may influence the observed comparative effectiveness of different types of reading programs with different types of students.

4. The present guidelines should be considered by decision-makers along with other pertinent and important information when they are making decisions about the types of reading programs to use at each grade level.

The guidelines are not to be viewed as rigid prescriptions, but as aids to decision-making.

5. The methods of the Evaluation System have been found to yield results that make the continuation and expansion of the Evaluation System warranted. Continuation and expansion can be expected to lead to more detailed and specific guidelines for choosing and implementing reading programs.

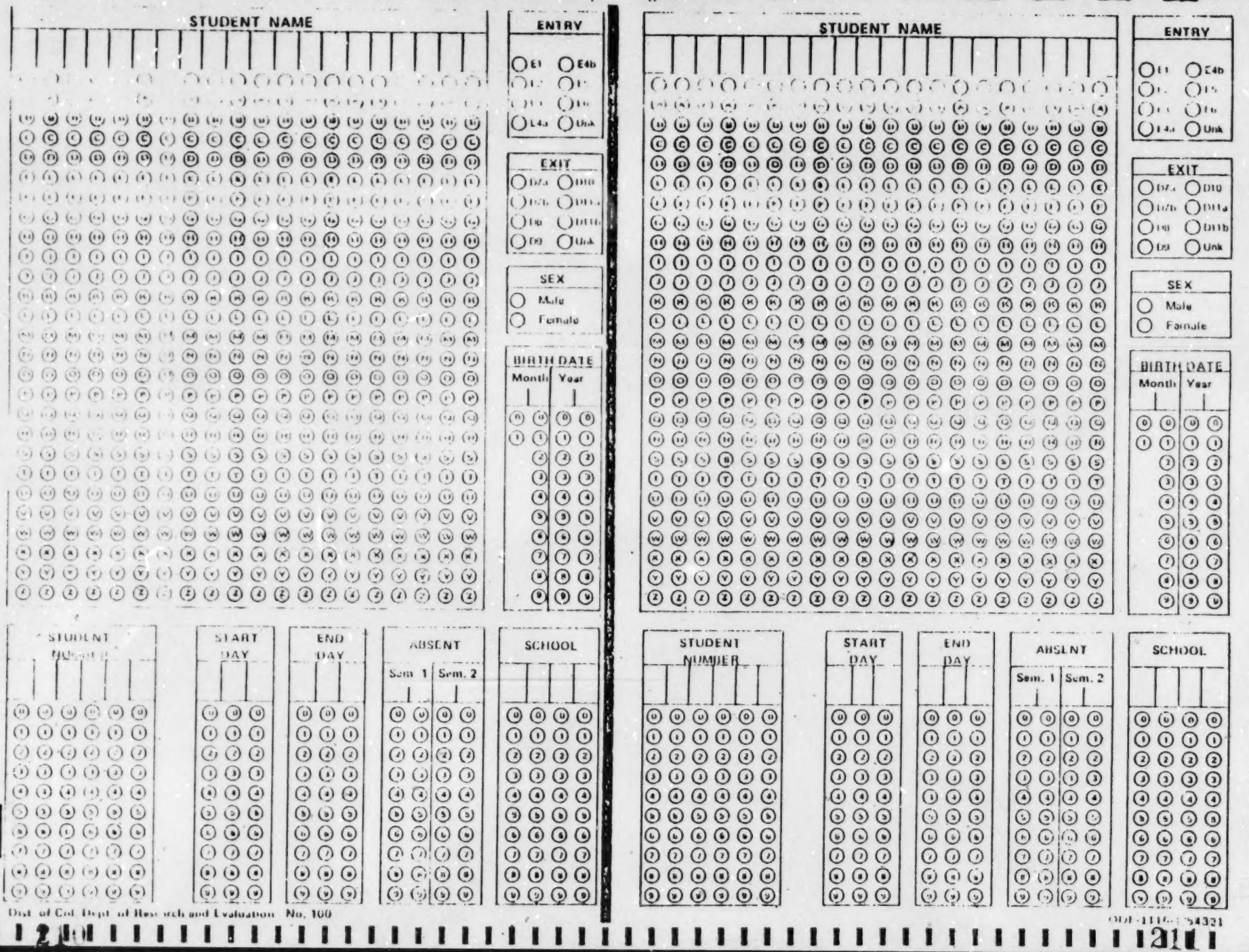
APPENDIX A



-191-

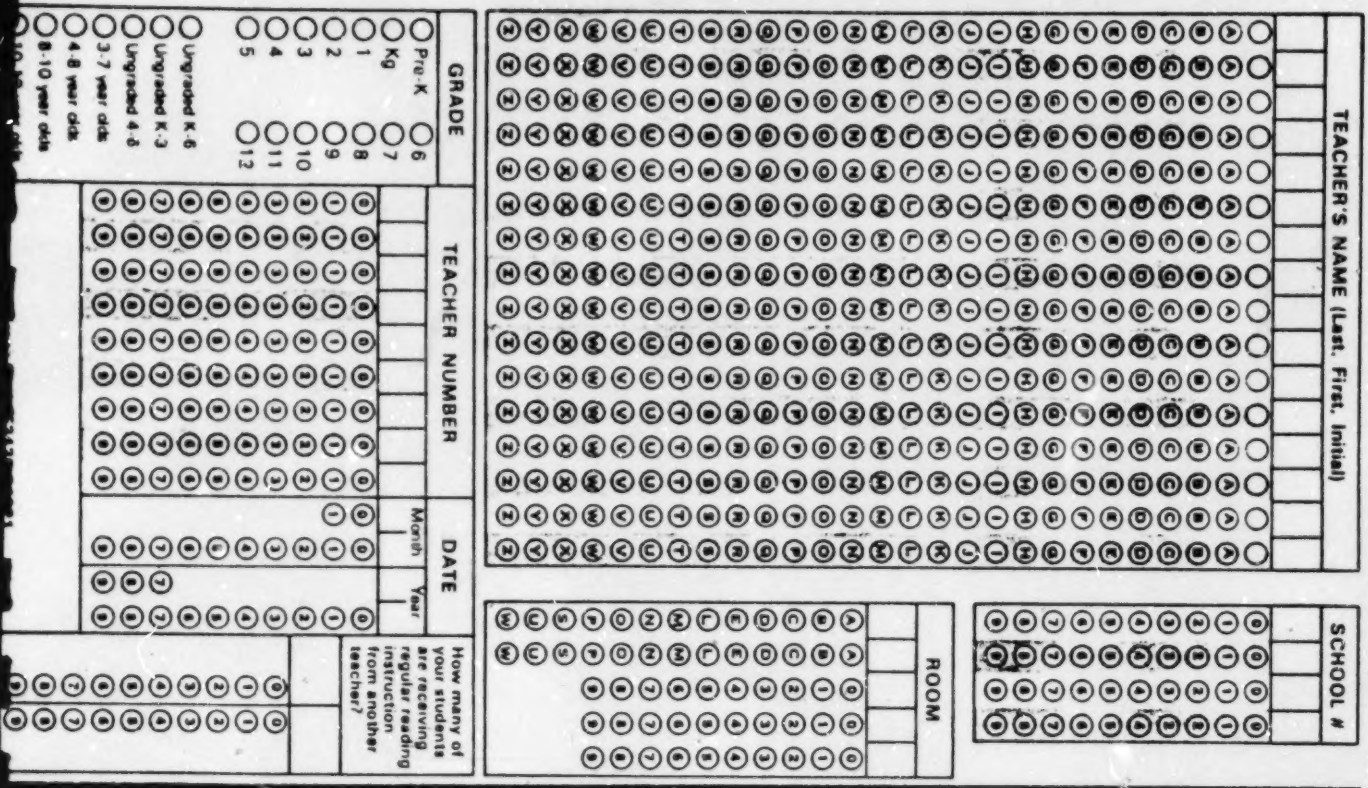
District of Col. Dept. of Research and Evaluation Form #101

NCS Trans Optic F3262-54321


STUDENT NAME
ENTRY
E1
E4b
E4a
Unk
EXIT
D10
D11b
D9
SEX
Male
Female
BIRTH DATE
Month
Year
STUDENT NUMBER
START DAY
END DAY
ABSENT
Sem. 1
Sem. 2
SCHOOL
Dist of Col Dept of Research and Evaluation No. 100

APPENDIX B

DISTRICT OF COLUMBIA DEPARTMENT OF RESEARCH AND EVALUATION FORM #201



The following questions pertain to your family or home-room class.

Often students have needs or problems that affect the teacher's instructional program or the students' ability to learn:

a. As a result, have you referred any of your students this year to any of the following?

- ○ Principal
- ○ Counselor
- ○ Health Specialist (Nurse)
- ○ Speech Therapist
- ○ Pupil Personnel Worker
- ○ Visiting Instructor
- ○ Special Teacher(s) (Reading, Math Specialists, etc.)
- ○ School Psychologist
- ○ Social Worker
- ○ Aides (Educational, Title I, Health, Community, etc.)
- ○ None
- ○ Other (Specify) _____

b. Please mark the number of students you referred this year because they had learning problems in one of the following categories: Reading; Math; Speech; Language Arts; Other

	1 - 3	4 - 6	7 - 9	10 or more
Reading	○	○	○	○
Math	○	○	○	○
Speech	○	○	○	○
Language Arts	○	○	○	○
Other	○	○	○	○

c. Please mark the number of students you referred this year because they had learning barriers in one of the following categories:

	1 - 3	4 - 6	7 - 9	10 or more
Behavior*	○	○	○	○
Attendance	○	○	○	○
Truancy	○	○	○	○
Nutrition	○	○	○	○
Drugs	○	○	○	○
Home or Community	○	○	○	○
Child Abuse	○	○	○	○
Economic Conditions (Clothing, shoes, etc.)	○	○	○	○
Medical/Physical Health (Vision, Sickness, Dental, Pregnancy, Injury, etc.)	○	○	○	○

d. (Answer only if you have made NO referrals) Why haven't you made referrals?

Learning Problems
Learning Barriers

- ○○ None of my students had a need to be referred
- ○○ I prefer to deal with the problems of my students myself if at all possible
- ○○ The necessary special teachers were not available
- ○○ There was no counselor available
- ○○ There was no health specialist available (nurse, health aide, etc.)
- ○○ There is a lack of response to referrals
- ○○ I lack confidence in our referral system
- ○○ I lack confidence in the resource personnel
- ○○ The referral procedures are poor
- ○○ We are encouraged by the school's policy to handle our own classroom problems
- ○○ Other

TRICT OF COLUMBIA DEPARTMENT OF RESEARCH AND EVALUATION FORM #202

THIS SIDE TO BE COMPLETED BY READING TEACHER

TEACHER NUMBER

SCHOOL NO

DATE Month Year

GRADE: Pre-K, Kg, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, Ungraded K-6, Ungraded K-3, Ungraded 4-6, 3-7 year olds, 4-8 year olds, 8-10 year olds, 10-12 year olds

TEACHER'S NAME (Last, First, Initial)

ROOM

How many regular reading classes do you teach during the school day?

- one
- two
- three
- four
- five
- six
- seven
- eight

If you teach only one reading class during the day, during what time of the day is it taught?

- before 10 am
- between 10 am and noon
- afternoon

How long have you been the reading teacher for your class or classes this year?

- less than 1 month
- 1 month
- 2-3 months
- 4-5 months
- 6-7 months
- 8-9 months

Have you been teaching your reading class or classes by a team teaching approach this year?

- yes
- no
- no, but we exchange students for reading

Have you taught any reading skills objectives for your class this year other than those on the attached list?

- yes
- no

If yes, what percent of the reading skills objectives you taught were not contained in the attached PRT listing?

- less than 10%
- 10% to 20%
- 21% to 30%
- 31% to 40%
- 41% to 50%
- 51% to 60%
- 61% to 70%
- 71% to 80%
- over 80%

7. Did you teach PRT skills objectives from more than one of the test levels this year?
   - yes
   - no

8. Were you free to choose the reading skills objectives taught in your class(es) this year?
   - totally
   - to a great extent
   - to some extent
   - no

9. To what extent do you coordinate your instructional reading program with others in the school who also work with any student in your reading class or classes? (Mark all those that apply)

OTHER PERSONS WHO HELP MY STUDENTS IN READING

totally	to a great extent	to some extent	not at all	
○	○	○	○	The child's family or homeroom teacher (if you are not that person)
○	○	○	○	The reading specialist(s)
○	○	○	○	The reading resource teacher(s)
○	○	○	○	The reading aide(s)
○	○	○	○	The reading tutor(s)
○	○	○	○	Others (Specify) _____

Dist. of Col. Dept. of Research and Evaluation #200

[redacted]

Teacher Name _____ School Name _____

STUDENT NAME

STUDENT ID NUMBER

TEACHER NUMBER

SCHOOL

Teacher

Aide

Section

If you are not this student's family or homeroom teacher, please print your name and the name of the aide who helps you in the reading class, and group or section of this student if you have more than one group.

B. 1. How many months has the child been in your class this year?

○ Less than one month ○ 1 month ○ 2 months ○ 3 months ○ 4 months
○ 5 months ○ 6 months ○ 7 months ○ 8 months

B. 2. Is this child in need of extra help in reading?

○ Yes, very much ○ Yes, somewhat ○ No

B. 3. THIS YEAR from what sources OTHER THAN YOU has this child received reading instruction DURING THE REGULAR SCHOOL DAY? Please mark all applicable.

○ Don't know - new student

○ None, no extra help ○ College student tutor ○ Other elem. sch. student tutor ○ Parent tutor
○ Reading specialist/reading resource teacher ○ Teacher aide tutor ○ Jr. high student tutor ○ Urban service corps tutor
○ Service group tutor ○ Sr. high student tutor ○ Other

B. 4. To the best of your knowledge how many months has this help been provided by the sources marked above?

○ Less than one month ○ 1 month ○ 2 months ○ 3 months ○ 4 months
○ 5 months ○ 6 months ○ 7 months ○ 8 months

B. 5. How often each week does this child receive reading instruction from the sources listed above? Mark most appropriate.

○ 5 Days ○ 4 Days ○ 3 Days ○ 2 Days ○ 1 Day ○ Less than once a week

B. 6. If this child needs extra help in reading and is not receiving it, please mark the main reason.

○ No reading specialist/reading resource teacher ○ Lack of instructional materials ○ Other (please explain) _____
○ Lack of tutorial services ○ Lack of teacher aides
○ Lack of diagnostic information

D. Why are you using the materials listed on the reverse side of this form for this student?

○ They are the materials I have found most effective in the past.
○ They are the only materials most readily available in my school.
○ They are materials suggested or recommended by the principal or supervisor.
○ They are keyed to the school system reading objectives and criterion-referenced tests.
○ They are materials found most suitable to this student's needs.
○ I learned about the materials in an in-service program/workshop.
○ I selected them on the basis of my own diagnostic test and observations.
○ I selected them on the basis of standardized test results.
○ They are the materials used as part of the special program in this building.
○ Other _____

DO NOT WRITE OUTSIDE BOX

C. Aside from reading instruction, how often does this student read on his own in school? (library books, comics, etc.)

○ Often ○ Sometimes ○ Never

E. How would you assess this child's interest in reading?

○ Very interested
○ Moderately interested
○ Seldom interested
○ Not interested
○ Actively dislikes reading

NCS Trans-Optic F3369 54321

EVALUATION SYSTEM
STUDENT READING PROGRAM FORM

MARKING INSTRUCTIONS
1. Use black lead pencil only (No. 2½ or softer.)
2. Do not use ink or ballpoint pens.
3. Make heavy black marks that fill the circle completely.
4. Erase completely any mark you wish to change.
5. Do not make any stray marks on this sheet.

BASAL SERIES
101 Bank Street Readers – MacMillan
102 Basic Curriculum Series – Lippincott (McLeod, McCracken & Walcutt)
103 Betts Basic Readers – American Book Co.
104 Bookmark Reading Program – Harcourt, World & Brace
105 Ginn Basic Readers, 100 Edition (Russell, et. al.)
106 Ginn Basic Readers, 360 Program (Clymer, et. al.)
107 Harper and Row Basic Readers (Van Roekel, et. al.)
108 Houghton Mifflin – Reading for Meaning (McKee)
110 Lyons and Carnahan – Curriculum Motivation Series (Fay)
111 MacMillan Reading Program (Harris & Clark)
112 Reading Caravans – D. C. Heath Co. (Witty and Freeland)
113 Reading Experience and Development (Read) – American Book Co.
114 Scott, Foresman – Bright Horizons Reading Program
115 Scott, Foresman – Curriculum Foundation Series
116 Scott, Foresman – Open Highways Program
117 Sheldon Basic Readers – Allyn & Bacon
118 Cambridge Book Co. – Reading Series
119 D. C. Heath – Communications
120 Ginn Basic Readers, 300 Edition
121 Holt, Rhinehart, Winston – Basic Reading System
122 Houghton Mifflin – Interaction
123 MacMillan – Series R
124 New Dimensions in Education – Alfa Series
125 Scott, Foresman – Reading Systems
199 Other (Please specify when gridding in box below)

LINGUISTIC/PHONIC
201 Basic Reading Series – SRA – (Pig in a Jig Books)
202 Lippincott – Basic Reading
203 Merrill Linguistic Readers (Fries, Wilson & Rudolph)
205 Open Court Foundation Program
206 Palo Alto Reading Program – Harcourt, Brace & World
207 Phonetic Keys to Reading Series – The Economy Co.
208 Phonics We Use – Lyons and Carnahan
209 Phono Visual – Phono Visual Products
211 Readiness in the Language Arts (RLA) – Cynthia Buchanan
299 Other (Please specify when gridding in box below)

PROGRAMMED/STRUCTURED
301 Lift-Off to Reading – SRA – (Woolman)
302 Programmed Reading – Webster Division of McGraw Hill – (Buchanan & Sullivan)
303 Project READ – Behavioral Research Labs – (Sullivan's Reading)
304 Starter/101 – (Silver Burdett (O'Keefe)
305 Random House – Structural Reading
306 Reading Readiness Tapes – Ideal School Supply
399 Other (Please specify when gridding in box below)

LANGUAGE EXPERIENCE
401 Experience Charts – Commercial
402 Experience Charts – Made by Teacher with Class
403 Language Experience Readers – Noble & Noble (Chandler Reading Program)
499 Other (Please specify when gridding in box below)

INDIVIDUALIZED READING
501 Extensive Use of Variety of Published Materials
502 Little Owl Series (Classroom Library) – Holt, Rinehart, Winston
503 MacMillan Reading Spectrum
505 SRA Reading Labs
506 Extensive Use of Variety of Non-Published Materials
599 Other (Please specify when gridding in box below)

SUPPLEMENTARY MATERIALS
605 Invitations to Personal Reading (Classroom Library) – Scott, Foresman
609 Reading Comprehension Skills Exercises (D. C. Schools)
610 Readers Digest Services, Inc.
611 Singer, Random House Literature Series
612 Torch Lighter Kits
613 Weekly Readers
614 Teacher Made Materials
615 Other (Please specify when gridding in box below)

SPECIAL
702 Distar – SRA (Englemann)
703 Frostig's Developmental Program in Visual Perception – Follett
704 I/T/A
705 Kindergarten Keys – Economy Co.
799 Other (Please specify when gridding in box below)

MOST USED MATERIAL	2nd MOST USED MATERIAL	3rd MOST USED MATERIAL
Approx. Level	Approx. Level	Approx. Level
○ RR ○ 3²	○ RR ○ 3²	○ RR ○ 3²
○ PP ○ 4	○ PP ○ 4	○ PP ○ 4
○ P ○ 5	○ P ○ 5	○ P ○ 5
○ 1¹ ○ 6	○ 1¹ ○ 6	○ 1¹ ○ 6
○ 1² ○ 7	○ 1² ○ 7	○ 1² ○ 7
○ 2¹ ○ 8	○ 2¹ ○ 8	○ 2¹ ○ 8
○ 2² ○ 9	○ 2² ○ 9	○ 2² ○ 9
○ 3¹ or above	○ 3¹ or above	○ 3¹ or above
OTHER (199, 299, ETC.) SPECIFY	OTHER (199, 299, ETC.) SPECIFY	OTHER (199, 299, ETC.) SPECIFY

A. To what extent is the reading program described above indicative of what was used all year for this child for reading instruction? (Exclude changes in reading level)

○ Don't know - student new to this class
○ Very representative or very similar
○ Somewhat representative or somewhat similar
○ Not at all representative or not at all similar

APPENDIX C

EVALUATION SYSTEM
LIST OF PARTICIPANT SCHOOLS SCHOOL YEAR 1977-78

Region I (N=7)	Beers
	Congress Heights
	Friendship
	Hendley
	McGogney
	Malcolm X
	Savoy
Region II (N=10)	Amidon
	Bancroft
	Bowen
	Hardy
	Hyde
	Key
	Lafayette
	Mann
	Stoddert
	Syphax
Region III (N=4)	Benning
	Davis
	Thomas
	Weatherless
Region IV (N=5)	Elow
	Goding
	Miner
	Walker-Jones
	Wilson, J.O.
Region V (N=4)	Bruce-Monroe
	Cleveland
	Meyer
	Petworth
Region VI (N=4)	Bunker Hill
	Noyes
	Shaed
	Whittier

Total Schools = 34

APPENDIX D

TABLE A

CORRELATION MATRIX ACROSS ALL-GRADES REGRESSION ON POSTTEST SCORES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Grade	1.00000	.41791	-.26933	-.38156	-.26522	-.38910	.19249	-.13801	.15165	.91094	.07263	.05625
(2) Test Level Count	.41791	1.00000	-.21364	-.19574	-.21164	-.20782	.20998	-.09857	.11065	.41400	.07012	.04936
(3) Raw Pretest	-.26933	-.21364	1.00000	.55799	.99935	.56022	-.04184	.04574	-.01360	-.24270	-.15897	-.15978
(4) Raw Posttest	-.38156	-.19574	.55799	1.00000	.56090	.99133	-.05479	.08857	-.00531	-.26744	-.17508	-.25094
(5) Arc SQRT Pretest	-.26522	-.21164	.99935	.56090	1.00000	.56333	-.04095	.04390	-.01515	-.23800	-.15886	-.15851
(6) Arc SQRT Posttest	-.38910	-.20782	.56022	.99133	.56333	1.00000	-.05142	.09166	-.00644	-.27026	-.16019	-.26156
(7) Class Size	.19249	.20998	-.04184	-.05479	-.04095	-.05142	1.00000	-.07154	.27288	.18083	.07697	.08497
(8) % Males	-.13801	-.09857	.04574	.08857	.04390	.09166	-.07154	1.00000	-.02777	-.17244	-.01237	-.09754
(9) Class Turnover	.15165	.11065	-.01360	-.00531	-.01515	-.00644	.27288	-.02777	1.00000	.18121	-.01299	-.03944
(10) Mean Material Level	.91094	.41400	-.24270	-.26744	-.23800	-.27026	.18083	-.17244	.18121	1.00000	-.02708	-.09535
(11) Variability Material Level	.07263	.07012	-.15897	-.17508	-.15886	-.16019	.07687	-.01237	-.01299	-.03708	1.00000	.15743
(12) Reading on Own	.05625	.04936	-.15978	-.25094	-.15851	-.26156	.08497	-.09754	-.03944	-.09535	.15743	1.00000
(13) Reading Interest	.07209	.03869	-.14095	-.27157	-.13962	-.28235	.06405	.08506	-.01228	-.09762	.21316	.23325
(14) Need for Help	.15311	.02197	.15908	.21529	.15953	.23353	.07605	-.10696	.21704	.36768	-.10908	.40441
(15) % Receiving Help	.19185	.06547	.00158	.03401	.00499	.04734	.08422	-.03244	.19779	.23678	-.03831	.05825
(16) Teacher Turnover	-.02293	.04073	.02356	-.04574	.02315	-.05062	-.01427	.03539	-.00805	-.06729	-.02551	.12647
(17) Teacher Absence	-.03015	.02753	-.04482	-.11284	-.04404	-.11753	-.04793	.03324	-.04368	-.05489	-.08111	.06626
(18) % Free Lunch	-.01887	.02847	-.18262	-.28347	-.18471	-.28753	-.04733	.03145	-.22720	-.18140	.21655	.28253
(19) Math Specialist	-.05629	-.09310	-.02732	-.01538	-.02647	-.00885	-.04148	-.06966	-.07445	-.06489	-.01927	-.05465
(20) Open Space	-.02033	.11434	-.00412	-.07998	-.00427	-.08645	-.02458	-.05890	.11537	.00403	.05126	-.06863
(21) Team Teaching	.04935	.07303	.07830	.08560	.08678	.07510	-.02091	-.05252	.11958	.11828	.03110	-.15918
(22) Reading Problems	-.13205	-.05767	-.07615	-.06551	-.07624	-.07545	-.02216	.09157	-.10236	-.23339	.09113	.13068
(23) Behavior Barriers	.06489	.06076	-.10057	-.08254	-.10100	-.08700	.01587	.05117	-.00139	.01615	.05199	.15087
(24) Health Barriers	-.04773	-.04105	-.06230	-.00042	-.06288	-.001	-.06875	.14297	-.01203	-.08166	.10487	.07321

	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(1) Grade	.07209	.15311	.19185	-.02293	-.03015	-.01887	-.05629	-.02033	.04935	-.13205	.06489	-.04773
(2) Test Level Count	.03869	.02197	.06547	.04073	.02753	.02847	-.09310	.11434	.07303	-.05767	.06076	-.04105
(3) Raw Pretest	-.14095	.15908	.00158	.02356	-.04482	-.18262	-.02732	-.00412	.07830	-.07615	-.10057	-.06230
(4) Raw Posttest	-.27157	.21529	.03401	-.04574	-.11284	-.28347	-.01538	-.07998	.08560	-.06551	-.08254	-.00042
(5) Arc SQRT Pretest	-.13962	.15953	.00499	.02315	-.04404	-.18471	-.02647	-.00427	.08678	-.07624	-.10100	-.06288
(6) Arc SQRT Posttest	-.28235	.23353	.04734	-.05062	-.11753	-.28753	-.00885	-.08645	.07510	-.07545	-.08700	-.00100
(7) Class Size	.06405	.07605	.08422	-.01427	-.04793	-.04733	-.04148	-.02458	-.02091	-.02216	.01587	-.06875
(8) % Males	-.08506	-.10696	-.03244	.03539	.03324	.03145	-.06966	-.05890	-.05252	.09157	.05117	.14297
(9) Class Turnover	-.03228	.21704	.19779	-.00805	.04368	-.22720	-.07445	.11537	.11958	-.10236	-.00139	-.01203
(10) Mean Material Level	-.09762	.36768	.23678	-.06729	-.05489	-.18140	-.06489	.00403	.11828	-.23339	-.01615	-.08166
(11) Variability Material Level	.21316	-.10908	-.03831	-.02551	-.08111	.21655	-.01927	.05126	.03110	.09113	-.05199	.10489
(12) Reading on Own	.23325	.40441	.05825	.12647	.06626	.28253	-.05465	-.06863	-.15918	.13068	-.15087	.05321
(13) Reading Interest	1.00000	-.23384	-.04725	.15750	.08417	.31114	-.02456	-.09697	-.14660	.13990	-.15993	.02999
(14) Need for Help	-.37384	1.00000	.26177	-.07059	-.07476	-.44644	-.03018	.00960	.25757	-.28375	.23543	-.21406
(15) % Receiving Help	-.04725	.26177	1.00000	-.04853	-.03393	-.22211	-.02650	.02055	.06272	-.20446	-.06407	-.04932
(16) Teacher Turnover	.15750	-.07059	-.04853	1.00000	-.08522	.06689	.01989	.02344	-.04938	.06024	-.02644	
(17) Teacher Absence	.08417	-.07476	-.03393	-.08522	1.00000	.08644	-.02681	.14615	-.00184	.09357	-.05713	.11791
(18) % Free Lunch	.31114	-.44644	-.22211	.06689	.08644	1.00000	.05845	-.02696	-.12591	.22075	-.11183	.16781
(19) Math Specialist	-.02456	-.03018	-.02650	.01989	-.02681	.05845	1.00000	-.03657	-.02447	-.02468	-.04029	-.24091
(20) Open Space	-.09697	.00960	.02055	.02344	.14615	-.02696	-.03657	1.00000	-.46274	-.12185	-.04895	-.04033
(21) Team Teaching	-.14660	.25757	.06272	-.04938	-.00184	-.12591	-.02447	-.46274	1.00000	-.15318	-.15114	-.22711
(22) Reading Problems	.13990	-.28375	-.20446	.06024	.09357	.22075	-.02468	-.12185	-.15318	1.00000	-.26150	.24084
(23) Behavior Barriers	-.15993	.23543	-.06407	-.02644	-.05713	-.11183	-.04895	-.04895	-.15114	-.26150	1.00000	-.26566
(24) Health Barriers	.02999	-.21466	-.04932	-.02644	.11791	.16781	-.24091	-.04033	-.22711	.24084	-.26566	1.00000

TABLE B

APPENDIX E

CORRELATION MATRIX PRIMARY GRADES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Grade	1.00000	.46247	-.46248	-.17618	-.14259	-.46082	.14862	-.11551	.04125	.78643	.27676	-.013
(2) Test Level Count	.46248	1.00000	-.17443	-.14259	-.11761	-.17361	.21232	-.15431	.08620	.49272	.11554	-.077
(3) Raw Pretest	-.46248	-.17443	1.00000	.52395	.43211	.99786	-.05850	.08597	-.07321	-.29755	-.10611	.174
(4) Raw Posttest	-.17618	-.14259	.52395	1.00000	.82693	.52558	-.05914	.07174	-.06272	-.29556	-.10432	.174
(5) Arc SQRT Pretest	-.14259	-.11761	.43211	.82693	1.00000	.43376	-.04876	.05912	-.05171	-.24373	-.08604	.143
(6) Arc SQRT Posttest	-.46082	-.17361	.99786	.52558	.43376	1.00000	-.05850	.07174	-.06272	-.29556	-.10432	.174
(7) Class Size	.14862	.21232	-.05850	-.05914	-.04876	-.05850	1.00000	-.07455	.02544	.03894	-.17167	-.349
(8) % Males	-.11551	-.15431	.08597	.07174	.05912	.07174	-.07455	1.00000	-.15343	.02705	.13943	.154
(9) Class Turnover	.04125	.08620	-.07321	-.06272	-.05171	-.06272	.02544	-.15343	1.00000	-.16685	-.01245	.025
(10) Mean Material Level	.78643	.49272	-.29755	-.29556	-.24373	-.29556	.03894	.02705	-.16685	1.00000	.09359	-.011
(11) Variability Material Level	.27676	.11554	-.10611	-.10432	-.08604	-.10432	-.17167	.13943	-.01245	.09359	1.00000	.249
(12) Reading on Own	-.01304	-.07748	.17435	.17422	.14376	.17422	-.34907	.15431	.02555	.19737	.24926	1.0000
(13) Reading Interest	.00208	-.06297	.10289	.08405	.06929	.08405	-.00016	-.00524	-.00524	.04926	.08697	.0990
(14) Need for Help	.09013	.13449	-.17222	-.17224	-.14241	-.17224	.42855	-.10772	.02544	-.23375	.13764	.8476
(15) % Receiving Help	.06255	.13461	-.02542	-.02545	-.01941	-.02545	.06580	-.03587	.01414	.02625	.03393	-.435
(16) Teacher Turnover	-.04970	-.01661	-.03602	-.03603	-.02771	-.03603	.61672	-.01328	.04662	.00273	.02214	-.132
(17) Teacher Absence	.02370	.02401	-.01894	-.01896	-.01461	-.01896	.32477	-.00699	.00246	-.00746	-.06150	.170
(18) % Free Lunch	-.03679	-.06404	.16076	.17141	.13214	.17224	-.05150	.03156	-.01954	-.02040	.09671	.084
(19) Math Specialist	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.000
(20) Open Space	-.03744	.19400	-.00236	-.00236	-.00182	-.00236	.13493	-.05246	-.00241	.15771	.07132	.270
(21) Team Teaching	-.03280	-.00347	.04721	.04547	.03511	.04721	.06233	-.05734	.07771	-.09765	-.03200	.074
(22) Reading Problems	-.08054	-.07678	-.05301	-.05395	-.04171	-.05301	-.12597	.01287	.11226	-.13136	-.23503	.017
(23) Behavior Barriers	.12810	.02853	-.08763	-.08924	-.06901	-.08763	.11671	-.04752	.15947	-.03473	-.05053	.102
(24) Health Barriers	-.00954	.04684	-.05648	-.05648	-.04369	-.05648	.00000	.00000	.00000	.00000	.00000	.000

	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(1) Grade	.00208	.09013	.06255	-.04970	.02370	-.03679	.00000	-.03744	-.03280	-.08054	.12810	-.009
(2) Test Level Count	-.06297	.13449	.13461	-.01661	.02401	-.06404	.00000	.19400	-.00347	-.07678	.02853	-.046
(3) Raw Pretest	.10289	-.17222	-.02542	-.03602	-.01894	.16076	.00000	-.00236	.04721	-.05301	-.08763	.056
(4) Raw Posttest	.08405	-.17224	-.02545	-.03603	-.01896	.17141	.00000	-.00236	.04547	-.05395	-.08924	.056
(5) Arc SQRT Pretest	.06929	-.14241	-.01941	-.02771	-.01461	.13214	.00000	-.00182	.03511	-.04171	-.06901	.043
(6) Arc SQRT Posttest	.08405	-.17224	-.02545	-.03603	-.01896	.17224	.00000	-.00236	.04721	-.05301	-.08763	.056
(7) Class Size	-.00016	.42855	.06580	.61672	.32477	-.05150	.00000	.13493	.06233	-.12597	.11671	.000
(8) % Males	-.00524	-.10772	-.03587	-.01328	-.00699	.03156	.00000	-.05246	-.05734	.01287	-.04752	.000
(9) Class Turnover	-.00524	.02544	.01414	.04662	.00246	-.01954	.00000	-.00241	.07771	.11226	.15947	.000
(10) Mean Material Level	.04926	-.23375	.02625	.00273	-.00746	.05273	.00000	.15771	-.09765	-.13136	-.03473	.000
(11) Variability Material Level	.08697	.13764	.03393	.02214	-.06150	.09671	.00000	.07132	-.03200	-.23503	-.05053	.000
(12) Reading on Own	.09900	.84760	-.43532	-.15229	.17000	-.08496	.00000	.27000	-.07400	.01766	.10299	.000
(13) Reading Interest	1.00000	-.40004	-.09036	.12577	-.09036	.08496	.00000	-.07401	-.07604	.21080	.13350	-.025
(14) Need for Help	-.40004	1.00000	.34151	-.14247	.11521	-.08496	.00000	-.07401	-.07604	.17442	.14398	-.022
(15) % Receiving Help	-.09036	.34151	1.00000	-.04004	.03179	-.02346	.00000	-.01549	-.01683	-.37187	-.26525	-.125
(16) Teacher Turnover	.12577	-.14247	-.04004	1.00000	.31221	-.02346	.00000	-.03040	-.03124	-.25747	-.06544	-.044
(17) Teacher Absence	-.09036	.11521	.03179	.31221	1.00000	-.01862	.00000	.03644	-.02484	.08387	.14603	.003
(18) % Free Lunch	.08496	-.08496	-.02346	-.02346	-.01862	1.00000	.00000	.14193	-.02533	.02374	.04742	.159
(19) Math Specialist	.00000	.00000	.00000	.00000	.00000	.00000	1.00000	.11321	-.11466	.32366	.06799	.182
(20) Open Space	-.03721	.01549	-.01683	-.03040	.03644	.14193	.11321	1.00000	-.04154	-.02136	-.08654	-.012
(21) Team Teaching	-.07089	.12233	-.00984	-.03124	-.02484	-.02533	-.11466	-.04154	1.00000	-.00617	-.08873	.018
(22) Reading Problems	.17442	-.37187	-.25749	-.25747	.08387	.02374	.32366	-.02136	-.00617	1.00000	.20762	.311
(23) Behavior Barriers	.14398	-.26525	-.06544	-.06544	.14603	.04742	.06799	-.08654	-.08873	.20762	1.00000	.191
(24) Health Barriers	-.02265	-.19256	-.04422	.00376	.00376	.15923	.18226	-.01242	.01817	.31152	.19115	1.000

APPENDIX F

TABLE C

CORRELATION MATRIX FOR INTERMEDIATE GRADES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Grade	1.00000	-.17473	.34598	-.[illegible]	-.[illegible]	-.[illegible]	.11715	-.01[illegible]	.02[illegible]	.67782	-.[illegible]	.03752
(2) Test Level Count	-.17473	1.00000	-.08732	.[illegible]	-.[illegible]	.[illegible]	.10654	.02[illegible]	.[illegible]	-.29953	.[illegible]	.12106
(3) Raw Pretest	.34598	-.08732	1.00000	.[illegible]	.[illegible]	.[illegible]	.05360	-.06651	.[illegible]	.37550	-.22467	-.12368
(4) Raw Posttest	-.08895	.04999	.50057	1.[illegible]	.[illegible]	.[illegible]	.01113	-.01[illegible]	.[illegible]	.09688	-.19319	-.15005
(5) Arc SQRT Pretest	.33884	-.06761	.79806	.[illegible]	1.[illegible]	.[illegible]	.03349	-.04755	.[illegible]	.37631	-.22016	-.12128
(6) Arc SQRT Posttest	-.07513	.[illegible]2458	.50429	.[illegible]	.51370	1.[illegible]	.[illegible]3232	-.01[illegible]	.[illegible]	.10382	-.13255	-.15935
(7) Class Size	.11735	.10654	.05260	.[illegible]1113	.[illegible]0544	.[illegible]	1.00000	.00[illegible]	.24110	.05595	-.00425	-.06379
(8) % Males	-.01005	.02169	-.06651	.01535	-.04755	.[illegible]1773	.04042	1.00000	-.06447	-.06124	-.01059	.17672
(9) Class Turnover	.05208	.05005	.03192	.08368	.03691	.[illegible]0769	.24110	-.04447	1.00000	.15404	-.13930	-.07415
(10) Mean Material Level	.65782	-.29953	.37530	.[illegible]9[illegible]	.[illegible]3[illegible]	.[illegible]3[illegible]	.05595	-.04124	.15404	1.00000	-.50425	-.24474
(11) Variability Material Level	-.00012	.28128	-.22027	-.[illegible]1[illegible]	-.22016	-.[illegible]1[illegible]	-.00425	-.01059	-.13930	-.50425	1.00000	.22618
(12) Reading on Own	.03752	.12106	-.12388	-.[illegible]1[illegible]	-.12125	-.[illegible]1[illegible]	-.04379	.17672	-.07415	-.24474	.22618	1.00000
(13) Reading Interest	.08368	.08681	-.15950	-.[illegible]1[illegible]	-.15573	-.[illegible]1[illegible]	-.04190	.20346	-.08374	-.42835	.42369	.81594
(14) Need for Help	-.01231	-.21088	.25198	.[illegible]1[illegible]	.25196	.[illegible]1[illegible]	.04238	-.07056	.16[illegible]	-.43552	.26412	.43024
(15) % Receiving Help	.19896	-.10800	.06280	-.[illegible]1[illegible]	.06[illegible]	-.[illegible]0[illegible]	.03969	-.06624	.10[illegible]	.19980	-.10055	.00084
(16) Teacher Turnover	.05[illegible]	.13065	-.08706	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.04906	-.00557	-.04767	-.06716	.12936	.07416
(17) Teacher Absence	-.06448	.06036	-.09566	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.05311	-.00512	.[illegible]	-.13720	.06410	.04722
(18) % Free Lunch	.06770	.14052	-.25701	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.04425	-.04331	-.27541	-.29089	.36472	.30005
(19) Math Specialist	.01670	-.09195	.01228	-.[illegible]0[illegible]	.01130	-.[illegible]0[illegible]	-.04622	-.11691	-.10133	-.62545	-.02633	-.07347
(20) Open Space	-.08455	.06736	-.00031	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.03327	-.12111	.04179	.03499	-.05722	-.12268
(21) Team Teaching	-.00270	.08704	-.16971	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.07731	-.12514	.13084	.22972	-.05027	-.25106
(22) Reading Problems	-.08868	.04387	-.17566	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.00052	.03[illegible]	-.02967	-.29969	.17736	.04913
(23) Behavior Barriers	.00460	.07057	-.10384	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.02354	-.00511	-.01[illegible]	-.14980	.11719	.11095
(24) Health Barriers	.05477	-.07830	-.11559	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.07562	.1[illegible]12	.04[illegible]	-.04922	.10945	.11518

	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(1) Grade	.08368	-.01231	-.19896	.[illegible]1[illegible]	-.[illegible]0[illegible]	.[illegible]0[illegible]	.01670	-.08455	-.00270	-.08868	.00460	.05477
(2) Test Level Count	.08681	-.21088	.10800	-.[illegible]1[illegible]	.[illegible]0[illegible]	.[illegible]1[illegible]	-.09195	.06736	.08704	.04387	.07057	-.07830
(3) Raw Pretest	-.15950	.25196	-.06280	.[illegible]0[illegible]	-.[illegible]0[illegible]	-.[illegible]1[illegible]	.01228	-.00031	-.16971	-.17566	-.10384	-.11559
(4) Raw Posttest	-.21911	.17668	-.01629	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.06028	-.07049	-.16509	-.14691	-.10058	-.07912
(5) Arc SQRT Pretest	-.15575	.25040	-.06705	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.[illegible]1[illegible]	.01130	-.07040	-.16037	-.17269	-.10366	-.11613
(6) Arc SQRT Posttest	-.21788	.18248	-.00992	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.[illegible]1[illegible]	-.05324	-.03315	-.15089	-.15090	-.09965	-.07315
(7) Class Size	-.06190	.04258	.03969	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.04622	-.03327	-.07731	.00052	-.02354	-.07562
(8) % Males	.20546	-.07032	-.06624	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.11691	-.12111	-.12514	.03686	-.00511	.10812
(9) Class Turnover	-.08372	.25498	.10580	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.[illegible]0[illegible]	-.10183	.04179	.13084	-.02967	-.01398	.04250
(10) Mean Material Level	-.28855	.43539	.10960	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.[illegible]1[illegible]	-.02545	.03489	.22972	-.29969	-.14980	-.04922
(11) Variability Material Level	.29329	-.26412	-.10055	.[illegible]1[illegible]	.[illegible]0[illegible]	.[illegible]0[illegible]	.02653	-.02724	-.05027	.17736	.11719	.10945
(12) Reading on Own	.81592	-.40024	.00084	-.[illegible]1[illegible]	.04722	.[illegible]0[illegible]	-.07347	-.12268	-.25106	.04913	.11095	.11518
(13) Reading Interest	1.00000	-.37429	-.09021	.[illegible]1[illegible]	.04787	.[illegible]0[illegible]	-.06116	-.14616	-.23135	.09430	.13014	.13695
(14) Need for Help	-.37429	1.00000	.13842	-.[illegible]1[illegible]	-.[illegible]0[illegible]	-.[illegible]1[illegible]	.15349	.37772	.22772	-.37776	-.21960	-.22744
(15) % Receiving Help	-.02021	.13842	1.00000	-.[illegible]1[illegible]	-.04722	-.14121	-.02307	.12444	.11054	-.34924	-.07534	-.03553
(16) Teacher Turnover	.15777	-.08718	-.06657	1.[illegible]0[illegible]	1.[illegible]0[illegible]	.[illegible]2154	-.01341	.00378	.04123	-.00406	-.04256	-.07352
(17) Teacher Absence	.04799	-.06615	-.06729	-.[illegible]1[illegible]	1.[illegible]0[illegible]	-.[illegible]3458	-.04000	.12042	.00556	.18527	.02173	.04200
(18) % Free Lunch	.36217	-.39858	-.14121	.[illegible]4[illegible]	-.[illegible]3458	1.[illegible]0[illegible]	-.00183	-.05254	-.11578	.15947	.16226	.14743
(19) Math Specialist	.04269	-.06118	-.02307	-.[illegible]1[illegible]	-.[illegible]0[illegible]	-.[illegible]0[illegible]	1.00000	.05433	.03[illegible]	-.04764	-.05489	-.38245
(20) Open Space	-.16616	.15349	.12492	-.[illegible]1[illegible]	-.[illegible]0[illegible]	-.[illegible]0[illegible]	.05433	1.00000	.48651	-.23629	-.00792	-.07401
(21) Team Teaching	-.23185	.37735	.11054	-.[illegible]1[illegible]	-.[illegible]0[illegible]	-.[illegible]1[illegible]	.48651	1.00000	-.31621	-.21703	.34006	-.05662
(22) Reading Problems	.09950	-.37776	-.34924	.[illegible]0[illegible]	.[illegible]1[illegible]	.[illegible]1[illegible]	-.04964	-.23629	-.31621	1.00000	.34006	.13140
(23) Behavior Barriers	.13019	-.21960	-.07534	-.[illegible]4[illegible]	.[illegible]2173	.[illegible]6226	-.05489	-.00792	.34006	.34006	1.00000	.36625
(24) Health Barriers	.10695	-.22744	-.03553	-.[illegible]7[illegible]	.[illegible]4200	.[illegible]4743	-.38245	-.07401	-.05662	.13140	.36625	1.00000

APPENDIX G

TABLE D

NUMBER OF STUDENTS AND CLASSES IN 34 EVALUATION SYSTEM SCHOOLS BY YEAR AND GRADE

School Year	Official October Membership	Total Number of Classes	Number of Classes in the Analysis Pool (1)	Usable Samples from the Analysis Pool								
				For Primary and Intermediate Grades (2)			For Individual Grades (3)						
				(1-3)	(4-6)	Total	1	2	3	4	5	6	Total
1975-76	16,170	627	512	108	121	229	71	55	71	70	74	59	400
1976-77	15,382	570	551	191	205	396	57	62	53	65	63	76	376
1977-78	14,318	562	548	244	214	458	81	92	87	78	74	77	489

(1) Classes wherein the average membership of students was less than 100 days were excluded from the analysis pool. Additionally, for a class to be included certain ID matching criteria had to be met.

(2) A class was not included if a student within the class was missing data on any of the selected variables. Combinations, or multi-level grade classes were included. Only students with pre-post matched data on the same level PRT were used.

(3) Classes composed of combinations, or multi-level grades were excluded from the by grade level analyses. Only students with pre-post matched data on the same level PRT were used.